



University of Campinas
Institute of Computing



Anderson Carlos Sousa e Santos

Improvements on Human Skin Segmentation in
Digital Images

Melhorias na Segmentação de Pele Humana em
Imagens Digitais

CAMPINAS
2015



University of Campinas
Institute of Computing



Anderson Carlos Sousa e Santos

Improvements on Human Skin Segmentation in Digital Images

Melhorias na Segmentação de Pele Humana em Imagens Digitais

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Computer Science.

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Supervisor/ Orientador: Prof. Dr. Hélio Pedrini

Este exemplar corresponde à versão final da Dissertação defendida por Anderson Carlos Sousa e Santos e orientada pelo Prof. Dr. Hélio Pedrini.

A handwritten signature in blue ink that reads "Hélio Pedrini".

Supervisor

CAMPINAS
2015

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Maria Fabiana Bezerra Muller - CRB 8/6162

Sa59i Santos, Anderson Carlos Sousa e, 1990-
Improvements on human skin segmentation in digital images / Anderson Carlos Sousa e Santos. – Campinas, SP : [s.n.], 2015.

Orientador: Hélio Pedrini.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Computação.

1. Visão por computador. 2. Processamento de imagens. 3. Segmentação de imagens. I. Pedrini, Hélio, 1963-. II. Universidade Estadual de Campinas. Instituto de Computação. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Melhorias na segmentação de pele humana em imagens digitais

Palavras-chave em inglês:

Computer vision

Image processing

Image segmentation

Área de concentração: Ciência da Computação

Titulação: Mestre em Ciência da Computação

Banca examinadora:

Hélio Pedrini [Orientador]

Moacir Pereira Ponti Junior

Neucimar Jerônimo Leite

Data de defesa: 08-06-2015

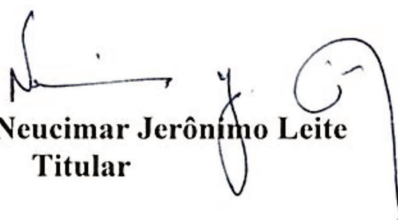
Programa de Pós-Graduação: Ciência da Computação

TERMO DE APROVAÇÃO

Defesa de Dissertação de Mestrado em Ciência da Computação, apresentada pelo(a) Mestrando(a) **Anderson Carlos Sousa e Santos**, aprovado(a) em **08 de julho de 2015**, pela Banca examinadora composta pelos Professores(as) Doutores(as):



Prof(a). Dr(a). Moacir Pereira Ponti Junior
Titular



Prof(a). Dr(a). Neucimar Jerônimo Leite
Titular



Prof(a). Dr(a). Hélio Pedrini
Presidente

Acknowledgements

I would like to thank my family for all the support during my whole life, for given me the strength to move forward and survive the obstacles. In special, my parents Machado and Conceição, my grandmother Jardelina and my sister Jamille for the love, affection and trust needed for accomplishing this.

I would like to thank Professor Hélio Pedrini for the opportunities and advices, for believing and trusting on my potential and for all time and patience dispensed in supervising this research.

I would like to thank the many friends conquered on my academic life, for all the shared moments of joy which made the journey unforgettable. A special thanks to Jacqueline and Lucas for all the care, concern and delicious lunches together.

I would like to thank my childhood friends, whom I can always count and who are on my side even at distance. Thank you for the happy moments and great memories.

I would like to thank Tilsa Isadora for the companionship, affection and attention, which made the path easier and more pleasant.

I would like to thank the faculty and staff members of the Institute of Computing at UNICAMP for the attention, operational support and precious learning, also for creating a healthy and inspiring environment.

An overwhelming gratitude to CNPq (process 152341/2014-5 from 2014 to 2015), CAPES (from 2013 to 2014) and FAPESP for the financial support.

Resumo

Segmentação de pele humana possui diversas aplicações nas áreas de visão computacional e reconhecimento de padrões, cujo propósito principal é distinguir regiões de pele e não pele em imagens. Apesar do elevado número de métodos disponíveis na literatura, a segmentação de pele com precisão ainda é uma tarefa desafiadora. Muitos métodos contam somente com a informação de cor, o que não discrimina completamente as regiões da imagem devido a variações nas condições de iluminação e à ambiguidade entre a cor da pele e do plano de fundo. Dessa forma, há ainda a demanda em melhorar a segmentação. Este trabalho apresenta três contribuições com respeito a essa necessidade. A primeira é um método autocontido para segmentação adaptativa de pele que faz uso de análise espacial para produzir regiões nas quais a cor da pele é estimada e, dessa forma, ajusta o padrão da cor para uma imagem em particular. A segunda é a introdução da detecção de saliência para, combinada com detectores de pele baseados em cor, realizar a remoção do plano de fundo, o que elimina muitas regiões de não pele. A terceira é uma melhoria baseada em textura utilizando superpixels para capturar energia de regiões na imagem filtrada, que é então utilizada para caracterizar regiões de não pele e assim eliminar a ambiguidade da cor adicionando um segundo voto. Resultados experimentais obtidos em bases de dados públicas comprovam uma melhoria significativa nos métodos propostos para segmentação de pele humana em comparação com abordagens disponíveis na literatura.

Abstract

Human skin segmentation has several applications on computer vision and pattern recognition fields, whose main purpose is to distinguish skin and non-skin regions. Despite the large number of methods available in the literature, accurate skin segmentation is still a challenging task. Many methods rely only on color information, which does not completely discriminate the image regions due to variations in lighting conditions and ambiguity between skin and background color. Therefore, there is still demand to improve the segmentation process. Three main contributions toward this need are presented in this work. The first is a self-contained method for adaptive skin segmentation that makes use of spatial analysis to produce regions from which the overall skin color can be estimated and such that the color model is adjusted to a particular image. The second is the combination of saliency detection with color skin segmentation, which performs a background removal to eliminate non-skin regions. The third is a texture-based improvement using superpixels to capture energy of regions in the filtered image, employed to characterize non-skin regions and thus eliminate color ambiguity adding a second vote. Experimental results on public data sets demonstrate a significant improvement of the proposed methods for human skin segmentation over state-of-the-art approaches.

List of Figures

2.1	Examples of some saliency detection maps.	18
2.2	Example of superpixel versus regular grid representation.	19
3.1	Main stages of the skin detection process.	23
3.2	Examples of images obtained by applying each stage of our method to an input image.	23
3.3	Examples of seeds detected through different thresholds for our method. . .	27
3.4	ROC curves for comparison of the tested methods.	28
3.5	Examples of skin detected with different methods.	30
4.1	Main stages of the proposed skin detection framework.	32
4.2	Examples of images obtained by applying each stage of our method to an input image.	33
4.3	ROC curves showing the results on ECU data set for the original method and the improvement with the framework.	34
4.4	ROC curves showing the results on IBTD data set for the original method and the improvement with the framework.	35
4.5	Comparative F_{score} results for original methods and our approach both in (a) ECU and (b) IBTD data sets.	36
4.6	Examples of skin regions detected through different methods.	37
5.1	Main stages of the proposed skin detection method.	39
5.2	ROC curve illustrating the results on test data set for the original method and the improvement through our method.	41
5.3	Examples of skin regions detected through different methods.	42

List of Tables

2.1	Statistics of ECU data set (Extracted from [44]).	21
3.1	Bin size evaluation for validation data set.	26
3.2	Evaluation of seed extraction for validation data set.	27
3.3	Detection results for different methods.	29
3.4	True positive rates for fixed values of false positive rate.	29
4.1	Detection results for different methods (ECU data set).	34
4.2	Detection results for different methods (IBTD data set).	35
4.3	True positive rates for a fixed value of false positive rate (ECU data set). .	36
4.4	True positive rates for a fixed value of false positive rate (IBTD data set). .	36
5.1	Detection results for different methods (ECU data set).	42
5.2	True positive rates for a fixed value of false positive rate (ECU data set). .	42

Contents

1	Introduction	11
1.1	Motivation	11
1.2	Objectives and Contributions	12
1.3	Organization	12
2	Background	13
2.1	Skin Segmentation	13
2.1.1	Pixel-Based Methods	13
2.1.2	Region-Based Methods	16
2.2	Saliency Detection	17
2.3	Supapixel Segmentation	19
2.4	Evaluation Metrics	20
2.5	Data Sets	21
3	A Self-Adaptation Method for Human Skin Segmentation Based on Seed Growing	22
3.1	Proposed Methodology	22
3.1.1	Seed Extraction	25
3.1.2	Propagation	25
3.1.3	Adaptive Detection	26
3.2	Experiments	26
3.3	Discussion	30
4	Human Skin Segmentation Improved by Saliency Detection	31
4.1	Proposed Methodology	31
4.2	Experiments	33
4.3	Discussion	36
5	Human Skin Segmentation Improved by Texture Energy Under Super-pixels	38
5.1	Proposed Methodology	38
5.2	Experiments	40
5.3	Discussion	43
6	Conclusions and Future Work	44

Chapter 1

Introduction

The human skin segmentation is the process in which its input is an image and the output corresponds to the pixels that belong to human skin, if any. This task can be perceived as a classification problem of two classes, skin and non-skin, where the main goal is to determine each class given the pixel features. In this work, the terms skin detection and skin segmentation will be used interchangeably.

1.1 Motivation

Human skin segmentation serves as a key step in a diverse of applications in image analysis. In this manner, if human skin is detected, there is a human, potentially the image contains a face and, if there is a large amount of visible skin for a single person, potentially there is nudity. From the segmented skin, it is possible to track the person's hand movements and then identify gestures. Summarily, the human skin segmentation aids human detection [11, 27], face detection [24, 25], nudity detection [46, 55, 53], gesture analysis [5, 66] and content-based image retrieval [9, 33] since it can benefit from any semantics extracted from the images.

However, the applications will still demand further processing and analysis, such that the skin detection configures as a preprocessing stage, which creates a requirement for speed and simplicity. The term simplicity here means that complex configuration, parameter selection and manual adaptation should be avoided. In other words, the method for skin segmentation should enable real-time applications and be adaptable to different purposes.

There are many challenges when working to segment skin in real-world images. First, the images acquired through different illumination conditions will present different skin characteristics. Furthermore, shadows, light intensity variation, reflections and person's pose will cause discrepancy between skin regions of the same image. Image resolution and size, as well as compression techniques, are also significant aspects.

Beyond the digital aspects, there are difficulties implied by the natural human skin characteristic: its color varies with ethnic diversity; its texture change with age — babies have a soft skin while elders have a more coarse skin; elasticity, which means that facial expressions and pose can change its aspect.

Although not robust to all these problems, color presents a meaningful evidence for skin detection. In fact, the vast majority of researches in the literature focus on color information to determine whether a pixel belongs to skin or not.

Nevertheless, there is an intrinsic problem associated with the use of color. It does not provide a separability between the pixels belonging to skin and one portion of non skin, referred to as skin-like pixels. This transition is considerably large and can frequently occur in an image, causing more errors than later applications could normally afford.

1.2 Objectives and Contributions

The main purpose of this work is to improve the human skin segmentation process through increasing the separability between skin and skin-like pixels. This dissertation presents three major contributions on this matter:

- A *self-adaptive method* that generates a skin color model specific to each image, which reduces color ambiguity and, that way, decreases the skin-like pixels. Although there are methods that perform adaptive segmentation, they usually rely on face detection or previous knowledge, whose both these issues can be hard to find or not be available. This method presents the novelty to be self-contained, that is, it uses spatial analysis to obtain true skin regions from which a specific color model is derived.
- A *saliency-based framework*, that uses saliency detection to remove background skin-like regions. A saliency detector captures the regions that “catch the eye” on an image. Thus, they can be used to separate foreground from background. A framework that combines color skin detection with a saliency detector is proposed to remove false positives from the first process.
- A *texture-based method* that models texture energy in skin and non-skin regions. It aims to remove the ambiguity caused by color, providing a second vote for classification. Texture is captured by a convolution filter and an energy measure is derived from each region to characterize it. The skin and non-skin energies are learned through Gaussian models, which are then applied to the image to obtain a skin texture probability. Skin texture and skin color are combined in a way that the pixel is considered skin only if both information agree.

1.3 Organization

This text is organized as follows. Chapter 2 presents relevant works related to skin segmentation, their advantages and drawbacks, the evolution of this field, as well as concepts and methods used in this dissertation. Chapter 3 details the self-adaptive proposed methodology and shows its results. Chapter 4 introduces the proposed saliency and color framework and presents its results. The texture-based improvement and its experimental results are shown in Chapter 5. Finally, Chapter 6 discusses the general contributions of the work and presents final conclusions and future directions.

Chapter 2

Background

This chapter describes some relevant works and concepts related to the skin segmentation problem in digital images.

2.1 Skin Segmentation

There are different approaches to skin segmentation and they can be categorized into pixel-based and region-based methods. In pixel-based approaches, the classification is performed by considering an individual pixel, independently on its neighbors. Region-based methods add contextual information by considering neighbors or the whole image.

This section briefly presents a review of human skin segmentation. For a more extensive survey, one could refer to Kawulok et al. [31].

2.1.1 Pixel-Based Methods

Initially, pixel-based methods are described and discussed for skin segmentation, which can be further categorized into rule-based, parametric models, non-parametric models, and adaptive.

Rule Based Methods

The simplest and older strategy for classifying a pixel is based on static decision rules that restrict skin to some positions on a chosen color space. Sobottaka et al. [56] proposed one of the first methods for skin detection using the HSV color space. They discarded the value channel and stated that the pixel will be considered as skin if $S \in [0.23, 0.68]$ and $H \in [0, 50]$. Later, Tsekeridou et al. [62] refined these rules to accommodate the value channel and add more conditions.

Many other methods with different rules and conditions have been developed since then [6, 7, 22, 23, 32, 34, 57, 68]. Soriano et al. [57] proposed a rule based on two quadratic functions for the normalized RG space. Kovac et al. [32] opted for the RGB space, however, with rules concerning the minimum and maximum values of the channel and their differences. Hsu et al. [23] adopted various thresholds that divide the HSI space on three zones that define the skin pixels.

An interesting approach was proposed by Cheddad et al. [7]. They perform a transformation from the RGB to one dimensional error signal e expressed as

$$e = (0.298936 \cdot R + 0.587043 \cdot G + 0.140209 \cdot B) - \max(G, B); \quad (2.1)$$

with RGB between $[0, 1]$. The pixel is considered as skin if $0.02511 \leq e \leq 0.1177$. This method achieves better results for general purpose human skin segmentation in comparison with many of other rule-based methods [52].

Parametric Models

Another approach is to fit a parametric model for the distribution of skin and non-skin color. Most of the methods available in the literature rely on mixture of Gaussians [27, 39, 61, 71], although there are also methods based on a single Gaussian [3, 59] or elliptical boundary model [37].

For a Gaussian Mixture Model, the skin or non-skin conditional probability is established as

$$P(c|class) = \sum_{i=1}^N w_i \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(c - \mu_i)^T \Sigma_i^{-1} (c - \mu_i)} \quad (2.2)$$

where c is the color vector, $class$ can be skin or non-skin and the contribution of the i -th Gaussian is determined by a scalar weight w_i , mean vector μ_i , and diagonal covariance matrix Σ_i . To train the models is used the Expectation-Maximization algorithm. More details can be found in [19].

Yang et al. [71] used the CIE LUV color space discarding the luminance and using 2 components on the mixture ($N = 2$). Terrillon et al. [61] compared nine different chrominance spaces using $N = 8$. McKenna et al. [39] used only Hue and Saturation to model a GMM of 2 components.

Jones et al. [27] used 16 Gaussian kernels and the RGB space to model both skin ($P(c|skin)$) and non-skin ($P(c|\neg skin)$) in order to define the probability of a pixel as skin given its color (c) to be

$$P(skin|c) = \frac{P(c|skin)P(skin)}{P(c|skin)P(skin) + P(c|\neg skin)P(\neg skin)} \quad (2.3)$$

as stated by the Bayes' rule. $P(skin)$ and $P(\neg skin)$ are the prior probabilities and usually are set to 0.5. Vezhnevets et al. [63] demonstrated that the choice of the prior probabilities does not influence the overall result.

With the parameters of the Gaussian Mixture that defines the skin probability, given an input image, every pixel is evaluated by Equations (2.2) and (2.3) resulting in a skin color probability map. Thus, the final segmentation can be performed through a threshold.

Non-Parametric Models

On the other hand, non-parametric models estimate the probabilities directly from the training data without any assumptions on its distribution shape. To do so, histograms

for skin ($H_{skin}(c)$) and non-skin ($H_{\neg skin}(c)$) are built over an annotated data set. The conditional probabilities are obtained as

$$P(c|skin) = \frac{H_{skin}(c)}{\sum(H_{skin}(i))} \quad (2.4)$$

$$P(c|\neg skin) = \frac{H_{\neg skin}(c)}{\sum(H_{\neg skin}(i))} \quad (2.5)$$

From these probabilities, the Bayes' rule in Equation (2.3) can be applied to produce the desired posterior probability. A skin color probability look-up table is then produced mapping every color value into a skin probability. In the test stage, given an input image, every color value is search on the look-up table to generate the skin color probability map.

Jones et al. [27] conducted an extensive evaluation of this method using the RGB space. A 3D histogram was built with different quantization values, ranging from 256^3 to 16^3 . They concluded that the bin size depends directly on the training set size, that is, some quantization is better than none. They also compared the approach with the Gaussian mixture model, which was found to perform better for a small training set, whereas the histogram model outperforms it for a sufficient large training data set.

Adaptive

There is a intrinsic overlap between skin and non-skin colors that limit the accuracy of color models [31]. Taking this fact into consideration, many researchers have adapted the previous mentioned methods according to the context. For instance, Kovac et al. [32] defined two rules: one for a uniform daylight illumination and other for flashlight lateral illumination. Still, this type of information requires a previous knowledge that can be unavailable.

Phung et al. [45] created an iterative method based on a homogeneity measurement for determining an optimal threshold, for the skin probability map of a particular image, by adapting it to maximize the homogeneous regions. In order to also adapt the threshold, Zhang et al. [73] used an Artificial Neural Network (ANN). Local minima are considered as candidates and 13 features are extracted from them. The neural network is trained to respond whether the candidate is an optimum threshold or not.

Lee et al. [36] also worked with ANN, however, to choose the best histogram model from a set of five models that were obtained by clustering models for individual training images. The input image is imposed to the neural network and, based on their responses, the best model is used in the detection. In the end, this method performs a lighting adaptation as the models represent the skin in different illumination conditions.

The most significant results are obtained by content-based adaptation, more specifically using face or hand detection. The first of such approaches [15] uses the region acquired by a face detector to update a unimodal Gaussian previously trained in the RG color space. To prevent non-skin facial pixels, such as hair and eyes to contribute to the skin model, it restricts the adaptation to pixels present in the original general model.

Stern et al. [58] used the facial region to select the best of five 2D color spaces (RG, rg, HS, YQ, C_bC_r). Each color space is evaluated in terms of their separability between

face pixels and non-skin surrounding pixels. With the appropriate color space, a Gaussian model is derived. Bilal et al. [4] used not only face, but also hand detection. To speed up the detection process, only C_b and C_r were used to define the histogram-based model.

Taylor et al. [60] have recently proposed to only use the face skin in RG to build a Gaussian model, discarding any previous training. They reduced the detected area to a circular region and used an outline detection technique to eliminate non-skin facial pixels.

A more robust technique is presented by Kawulok [29], which uses an eye detection to generate a facial region. All the pixels in this region are used to build a local skin histogram, whereas general skin and non skin histograms are built by using training data. It is reported that the histogram quantization should be smaller for the local model, and then extrapolate to the quantization of the general histogram. A $P_{face}(skin|c)$ is then derived through Bayes' Equation (2.3) using the local skin and general non skin models. Finally, the global and local probability maps are combined by weighted mean to produce a final skin probability map.

2.1.2 Region-Based Methods

Region-based methods are generally used as a second step for pixel-based methods. They add neighborhood information to increase the separability between skin and skin-like pixels. Wang et al. [67] used fixed rules for RGB and YC_bC_r , combined the result of both and used the grey-level co-occurrence matrix (GLCM) [17, 43] to extract textural features and classify the found skin regions. Although the false positive rate decreased, the true positives also decreased.

Ng et al. [42] applied the entropy of 2D Daubechies wavelets [17, 43] over a GMM skin color classification results. They clustered the texture vectors, using K-means, and eliminates the pixels from clusters whose mean is maximal, assuming that it will have rougher texture. This method is much dependent on the number of clusters and clusters to be eliminated. The improvement is not significant since the decrease in true skin detection is close to the false skin.

The histogram based skin probability map is used by Jiang et al. [26] to find initial skin candidates. Then, a lower threshold is used as a texture second step to eliminate the high number of skin-like pixels. Gabor wavelets are used to extract textural features that are combined to produce an untrained texture map. Therefore, a threshold on this map is required to eliminate non-skin texture. Similarly to other methods, this approach also disturb the true skin detection. Then, the authors used color and texture information to select markers of watershed segmentation [17, 43] to grow skin regions.

Similar strategies using region growing have been developed, which can be further categorized as spatial analysis [31]. They consider the structural alignment in the neighborhood of pixels classified as skin — generally with a probability map — such that it refines the segmentation process by removing false positives.

These methods apply region growing algorithms, where the seeds are found through a high threshold on a skin probability map. Ruiz et al. [48] proposed a diffusion process where a criterion considers the Euclidean distance in color space or skin probability between the seed pixel and its adjacent neighbors. A threshold decides if a neighbor must be

merged as skin. Additionally, there is a minimum threshold for the neighbor probability. The method is similar to a threshold hysteresis [17, 43], just adding a distance constraint.

An energy accumulation criterion was proposed by Kawulok [29]. The seed pixels receive energy 1 and pass an amount of energy to its neighbors based on their skin probabilities. The process continues until there is no more energy to be passed. This is done to avoid that the growing “leaks” to non-skin regions by imposing a limit to it.

Kawulok [30] also proposed a much more complex method that uses a distance transform and relies not only on skin probability but also on hue and luminance information. In order to perform the propagation, shortest routes from the seeds are calculated using Dijkstra’s algorithm [12]. The cost of a pixel x to adjoin a pixel y is defined as

$$\rho(x \rightarrow y) = \rho_I(x, y) \cdot [1 + \rho_P(x \rightarrow y)] \quad (2.6)$$

$$\rho_I(x, y) = \alpha_{diag} \cdot (|Y(x) - Y(y)| + |H(x) - H(y)|) \quad (2.7)$$

$$\rho_P(x \rightarrow y) = \begin{cases} \frac{P_t - P(y)}{1 - P_t} & \text{for } P(y) > P_\beta \\ \infty & \text{for } P(y) \leq P_\beta \end{cases} \quad (2.8)$$

where $Y(\cdot)$ is the luminance and $H(\cdot)$ is the hue obtained from the HSV color space, α_{diag} represents a penalty of $\sqrt{2}$ if the neighbor is at diagonal direction, $P(\cdot)$ is the probability, and P_t is a probability threshold that is said to be fixed in 0.6 and is not very sensitive in the results.

2.2 Saliency Detection

The goal of the saliency detection methods is to find the content in images which attracts human attention. There are two approaches toward this issue: estimation of the eye fixation points or determination of the saliency object. Here, we discuss the second approach in more details since it is coherent with our background elimination purpose.

The methods that perform figure-ground separation – a term to define the separation between object and background content – can be further classified based on the its domain: spatial frequency or color space.

The first relies on the signal within an image transform. Hou et al. [21] proposed the use of spectral residual – the difference between the log amplitude signal obtained with Fourier Transform (FT) and its smoothed version. Later, Guo et al. [18] showed that the image phase spectrum of the FT can generate the saliency map. More recently, Hou et al. [20] described an image signature based on Discrete Cosine Transform (DCT), which is used to obtain a saliency map that was proved to concentrate energy of spatially sparse foreground. This method is simple and provides good accuracy in terms of biological principles, however, we found that the produced map is too blurry (Figure 2.1b), which affects segmentation precision and also tends to work well only for small centered objects.

The methods that rely on color spatial domain can be further divided into using local or global contrast. They deal with the rarity of the foreground in relation to the vicinity of the pixel/region [1] or the entire image, respectively [8]. Local approaches are more sensitive to edges and noise, whereas global ones explore the overall structure and

relations, making them more suitable for segmentation. A combination of both [16] can also be beneficial.

The intuitive strategy of the global methods is to calculate the contrast of each region to the entire image. Nevertheless, this produces unsatisfied results and often classifies background regions as salient. Recent works have introduced boundary priors with the assumption that the image boundaries belong to the background [70] or the regions that most easily connect with the boundary to form the background [69].

The drawback of these methods is that they fail if the foreground object touches the boundary. In order to overcome this, Zhu et al. [75] proposed a measure of boundary connectivity, $BndCon$, defined as

$$BndCon(R) = \frac{|\{patch | patch \in R, patch \in Bnd\}|}{\sqrt{|\{patch | patch \in R\}|}} \quad (2.9)$$

where R is an image region, Bnd is the set of image boundary patches and $patch$ is a small image block. Regions that are heavily connected to the boundary, in other words, regions that have more patches belonging to the boundary set in relation to the total patches will produce larger values of boundary connectivity. For effective computation, they extend this notion through superpixels. Zhu et al. [75] also proposed to determine saliency with the minimization of a cost function that considers a background weight, a foreground weight and a smoothness term.

Figure 2.1 shows saliency map results for some of the described methods.

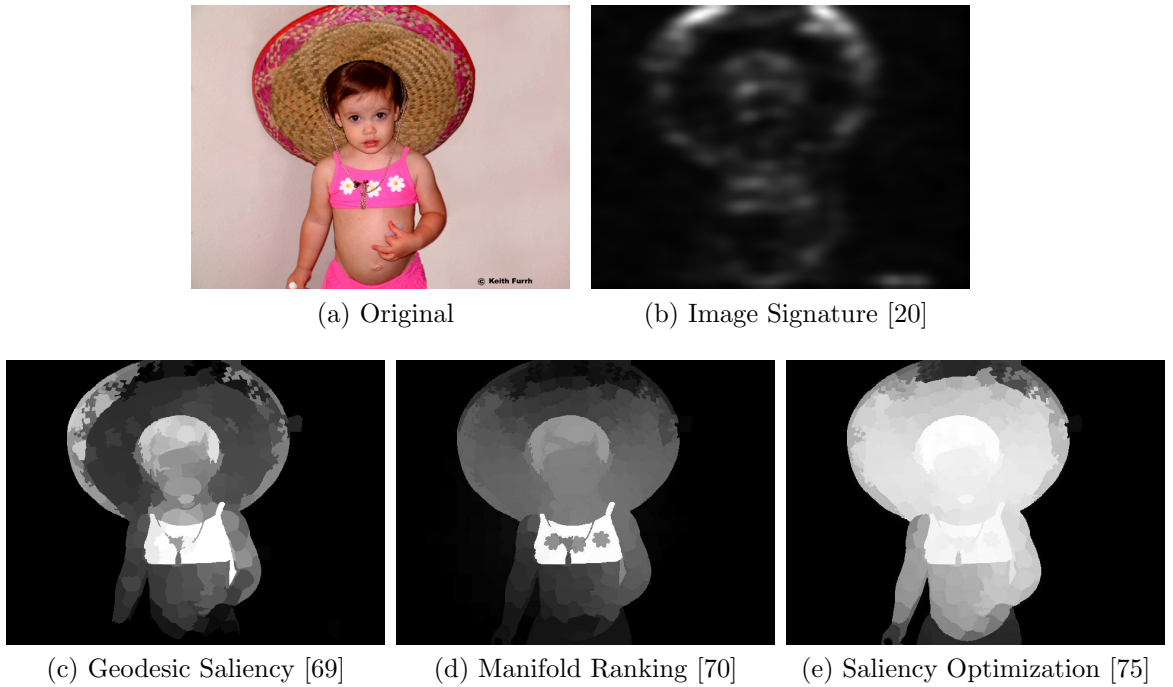


Figure 2.1: Examples of some saliency detection maps.

2.3 Superpixel Segmentation

In image analysis applications, there is often demand to work on a region rather than an individual pixel. Detection of objects, faces, people and other relevant contents are commonly done through a regular grid. Features are extracted for all windows of fixed size and verified to match the desired content. However, this is neither a good or natural representation. Inside a regular grid, there might be pixels that belong to different perceptually regions, raising noise and uncertainty.

A superpixel is an atomic region, formed by groups of perceptually meaningful pixels. It suppresses the need for a more reliable, unambiguous underlying representation. It also offers speed advantages for pixel-wise algorithms when the pixels are redundant in the information they provide. That way, superpixel segmentation algorithms are evaluated in their speed and adherence to boundaries. Figure 2.2 shows an example of an image divided into superpixels and regular grid.

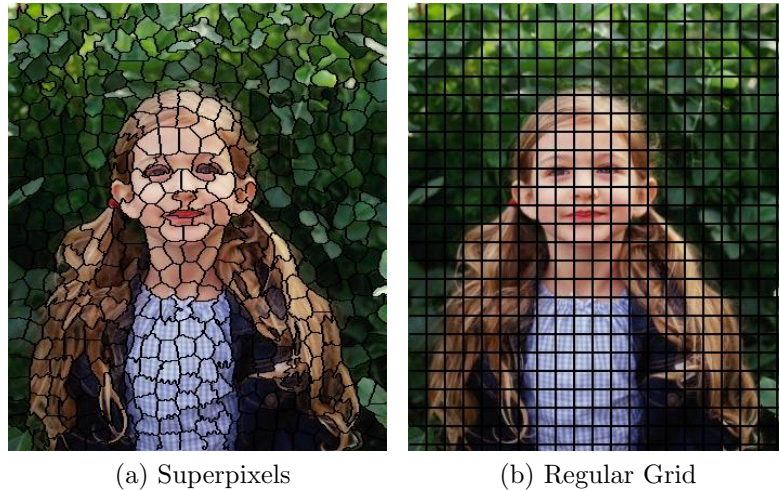


Figure 2.2: Example of superpixel versus regular grid representation.

Several methods have been developed to produce superpixels. Mori et al. [41] for instance, used a classical segmentation algorithm, namely Normalized Cuts [54], to obtain superpixels through an over-segmentation. Comaniciu et al. [10] used mean shift, an iterative general segmentation algorithm that performs clustering.

Other works [13, 14, 47, 64] followed the idea of using general purpose segmentation algorithms with an over segmentation criterion. Some problems that arise with those approaches include the fact that size, quantity and regularity of superpixels can be uncontrolled and the computational complexity may not be suitable for the preprocessing goal.

Moore et al. [40] proposed a segmentation method specific for generating superpixels. It uses a graph-based segmentation combined with a regular grid to conform the superpixels into a consistent topology, in order to hold the regularity and relationship with its neighbors.

Levinshtein et al. [38] proposed a geometric-flow based method that dilates seeds adapting to a local image structure. This method constraints the superpixels to uniform

sizes and compactness, however, it is still slow and has a poor boundary adherence.

To address common problems found in superpixel segmentation methods, Achanta et al. [2] proposed a fast, memory efficient method, named SLIC (Simple Linear Iterative Clustering). It is an adaptation of the well-known K-means algorithm that performs clustering by iteratively setting the elements to clusters with the nearest mean and updating their means.

SLIC implements two crucial changes on K-means: a spatial restriction to the distance calculations and the distance measure. It initially samples the clusters centers into a regular grid of $S \times S$, then calculates the distances of each center pixel C_k to every pixel i in a $2S \times 2S$ region around it. The pixels are assigned to the closest center and the center is updated to correspond to the new mean of the cluster.

The distance calculations are an important ingredient. They use the CIELAB color space and the x and y pixel positions to compose a 5D pixel representation: $[l \ a \ b \ x \ y]^T$. However, a 5D Euclidean distance is not suitable, as the position range of values varies with the grid size. Thus, they proposed to calculate color and space separately, both normalized and combined in the end. Equations (2.10), (2.11) and (2.12), taken from [2], show this process.

$$d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2} \quad (2.10)$$

$$d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (2.11)$$

$$D = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad (2.12)$$

where m is a constant and S is the maximum expected spacial distance. Notice that m can be used for balance between a regular size and shape and adherence to boundaries.

SLIC is the most commonly used method for superpixel segmentation since it is fast, provides high segmentation quality and is simple, requiring only the amount of superpixels in its default version.

2.4 Evaluation Metrics

The performance of the skin detection methods is measured through a number of metrics:

- True positive rate (η_{tp} - percentage of skin correctly classified as skin);
- False positive rate (δ_{fp} - percentage of non-skin classified as skin);
- Precision (η_{prec} - percentage of correctly classified pixels out of all the pixels classified as skin);
- F_{score} (harmonic mean between η_{prec} and η_{tp});
- Detection error ($\delta_{min} = (1 - \eta_{tp}) + \delta_{fp}$).

Additionally for probabilistic classification, the ROC (receiver operating characteristics) and the respective area under curve (AUC) are applied. In this dissertation we show the ROC curve relating true positives and false positives. That way the better curve is that with the largest area.

2.5 Data Sets

Three publicly available skin detection databases were used on this research. All of them provide a manually annotated ground-truth that makes possible to identify the pixel class (skin or non-skin) for the training and quantitatively evaluate the detection output.

The Compaq Database [27] contains images acquired from the Internet in a diverse variety of settings. There are 8963 non-skin images and 4666 skin images, which sum approximately 1 billion pixels. It is one of the largest databases for skin segmentation and is extensively used in the literature. The smallest image that contains skin has 38×39 pixels, the largest one has 1068×848 pixels, whereas the median 254×266 pixels.

The ECU Database [44] contains 4000 images, where most of them were collected manually from the Internet and a very small portion captured by researchers. In total, there are nearly 138 billion pixels, as the image sizes vary from 117×99 pixels to 1500×2000 pixels with median 584×464 pixels. They ensure a diversity in terms of background scenes, lighting conditions and skin types. Table 2.1, extracted from [44], shows the different types of skin and lighting conditions for ECU Database.

Table 2.1: Statistics of ECU data set (Extracted from [44]).

Skin Types	Images
whitish, pinkish	1665
yellowish, light brownish	1402
reddish, darkish, dark brownish	965
other skin types	102
Lighting Conditions	Images
indoor lighting conditions	1931
outdoor lighting conditions	1855
other lighting conditions	214

The IBTD Database [74] is composed of 555 images extracted from a larger database. The process occurred in the following way: 16,500 images with a minimum size of 150×150 were randomly selected from news, sports and entertainment websites available in the Internet, then these images were submitted to a filtering tool for objectionable images, where the suspicious images were retained in the data set. This means that these images present more skin-like pixels, which constitutes a more difficult database.

Chapter 3

A Self-Adaptation Method for Human Skin Segmentation Based on Seed Growing

Most studies [28, 31, 44] use color information as evidence for detecting skin since this property is able to provide computational efficiency while it demonstrates to be robust to occlusions and rotation and scaling transformations [28].

The main obstacle is the existence of an overlap between skin and non-skin colors that occurs independently of the color space, so objects made of wood, a wall or clothes are very often mistaken as skin. To minimize that, some segmentation methods [31] attempt to perform a color separability dependent on the skin color that appears in the image. The most significant improvements are performed with the aid of face detection, which provides an estimation of the remaining skin.

This chapter presents an adaptive segmentation method with no need for face or any other body content detection. It is based on an estimation from regions found through spatial analysis performed with a skin probability map. Experiments conducted on a large well known data set show that our method outperforms other skin segmentation approaches available in the literature.

The methods developed herein resulted in a paper [49] presented at the 10th International Conference on Computer Vision Theory and Applications (VISAPP'2015).

3.1 Proposed Methodology

We propose a method for skin segmentation that combines spatial analysis and adaptive models for better skin probability estimation. The methodology can be divided into three main steps. First, seeds are extracted from the general probability map through a precise and systematic strategy for spreading them over the images. Second, a controlled propagation method is applied to grow the seeds into skin blobs. Finally, these blobs are used to estimate the skin color present in the images and such that can optimize the probability map. The main steps of our method are presented in the diagram shown in Figure 3.1.

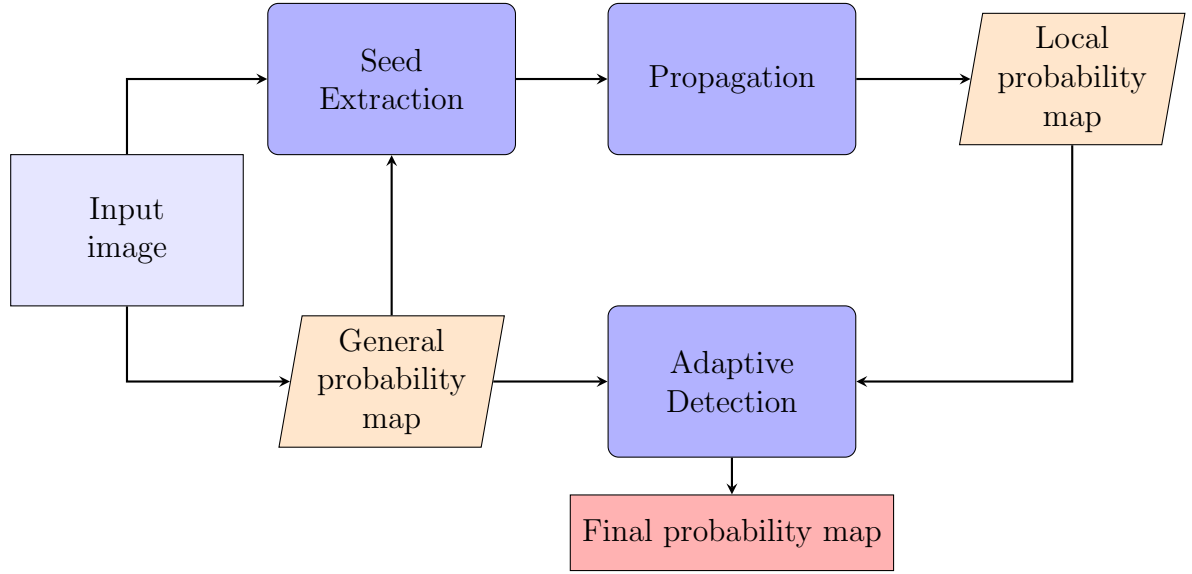


Figure 3.1: Main stages of the skin detection process.

Figure 3.2 illustrates the application of the proposed method to an input image, where the general probability map, the extracted seeds, the blobs after propagation, the final probability map and the resulting segmentation are shown.

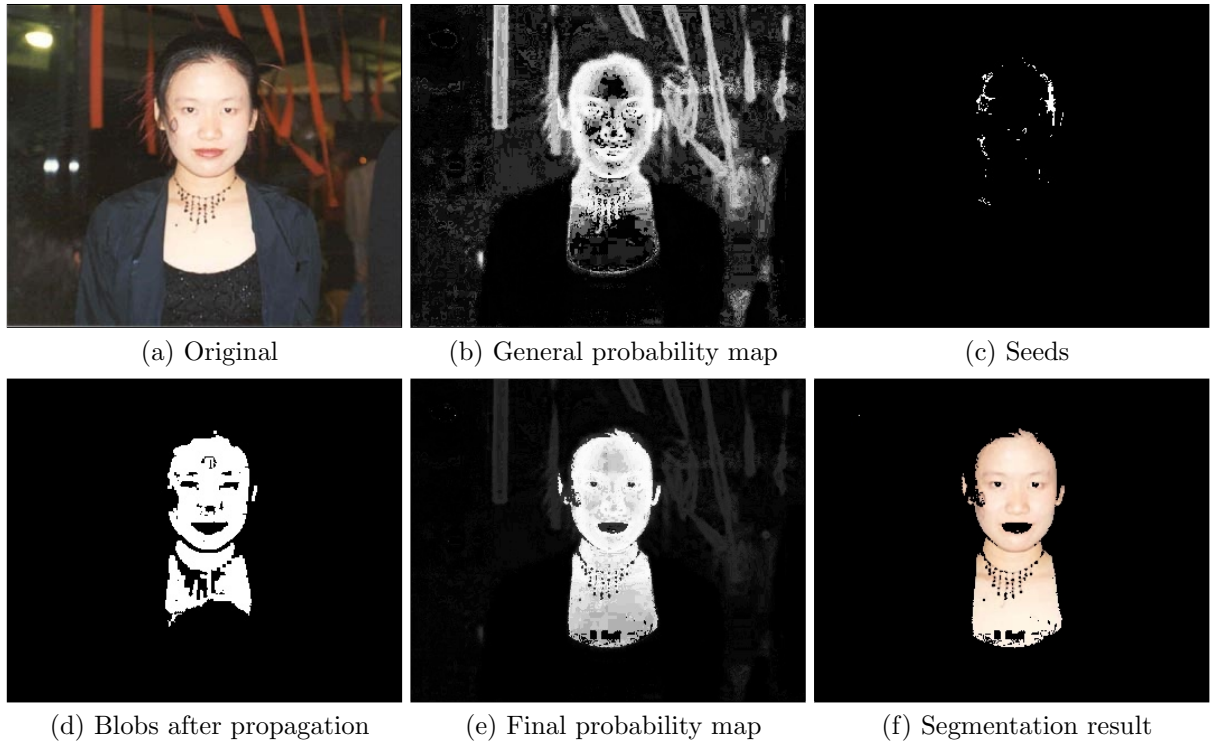


Figure 3.2: Examples of images obtained by applying each stage of our method to an input image.

Algorithm 1 describes in details the main steps performed by the proposed skin segmentation method.

Algorithm 1: Proposed skin segmentation method based on seed growing.

input : color image I , histogram of skin (H_{skin}) and non-skin ($H_{nonskin}$) colors.
output: Final probability map M_{final}

- 1 Build general probability map (M_{global}) according to Equations (2.3), (2.4), (2.5), using H_{skin} and $H_{nonskin}$
- 2 $M_{blur} \leftarrow \text{blur}(M_{global}, size)$
- 3 $T_{seed} \leftarrow \max(M_{blur})$
- 4 **if** $T_{seed} \leq T_{seed_{min}}$ **then**
- 5 | **return** M
- 6 **end**
- 7 $edges \leftarrow \text{edgeDetector}(I)$
- 8 **for** $x \in I$ **do**
- 9 | **if** $M_{global}(x) \geq T_{seed} \wedge x \notin edges$ **then**
- 10 | | $Seeds \leftarrow x$
- 11 | **end**
- 12 **end**
- 13 $Q \leftarrow \{Seeds\}$ {where Q is a priority queue}
- 14 **for** $x \in I$ **do**
- 15 | **if** $x \in Seeds$ **then**
- 16 | | $C(x) = 0$
- 17 | **else**
- 18 | | $C(x) = -1$
- 19 | **end**
- 20 **end**
- 21 **while** $Q \neq \emptyset$ **do**
- 22 | $q = \text{pop}(Q)$
- 23 | **for** $s \in \text{Neighbors}(q)$ **do** {8 - neighborhood}
- 24 | | $c = C(q) + \rho(q \rightarrow s)$
- 25 | | **if** $(c < C(s) \vee C(s) < 0) \wedge s \notin edges$ **then**
- 26 | | | $C(s) = c$
- 27 | | | $Q \leftarrow s$
- 28 | | **end**
- 29 | **end**
- 30 **end**
- 31 Normalize C by scaling the costs from 0 for the maximal cost to 1 for a zero cost
- 32 **for** $x \in I$ **do**
- 33 | **if** $C(x) > 0$ **then**
- 34 | | $H_{skin_local}(\text{color}(x))++$
- 35 | **end**
- 36 **end**
- 37 Build final probability map (M_{final}) according to Equation (3.2) using H_{skin_local}
- 38 **return** M_{final}

3.1.1 Seed Extraction

The most important step in a region growing algorithm is the proper choice of the seeds. Since seeds can correspond to false positives, the propagation process does not guarantee that inadequate seeds do not occur in the images, which can compromise the final estimation.

As described in Section 2.1, skin region propagation methods usually rely on a fixed high threshold and size based analysis for producing the seeds. Such methods do not take different characteristics of each image into account, as well as its respective probability map, once the same threshold is applied to all images. Another disadvantage is the assumption that the resulting skin-like seeds (false positives) are very small, which is not always true. Taking these factors into consideration, we propose an adaptive seed extraction with a homogeneity-based analysis.

In order to obtain the best high threshold for a particular image, we first apply a mean filter [17] to the probability map and then take the maximum probability (Line 3, Algorithm 1). To allow for images with no skin, if the maximum value is smaller than a minimum threshold ($T_{seed_{min}}$), it is discarded, otherwise it is assigned as the seed threshold for the original probability map. Therefore, we obtain seeds with high probability by considering the context.

To prevent the occurrence of false positives, we exclude the choice of seeds located in edge regions, since skin is usually a smooth and homogeneous region. The used edge detector is described in the following section.

3.1.2 Propagation

The objective here is to expand the seeds into skin blobs. The main drawback of propagation methods is the “leakages”, that is, the seed growth to a region of non skin. In order to avoid those, we establish a strict control of the propagation. We modified the cost propagation proposed in Kawulok [30] by adding a constraint in which the propagation cannot flow out the image edges. As a consequence, an increase in the false negative rate is expected with the reduction of the false positive rate. The next step will address these undetected skin regions.

Prior to the actual propagation, an edge detector is applied to the images. In order to benefit from color information, we utilize a color edge detection technique that combines (through logical *or* operator) the results of Canny detector [17] for each of the three channels in HSV color space. Following that, a morphological dilation operation is performed, such that small gaps can be closed.

The propagation process is similar to the work by Kawulok [30] except that the process stops when an edge is reached for a certain direction (Line 25, Algorithm 1). Besides preventing false positives, this also speeds up the algorithm, once the original approach calculates the costs from the seeds to every other pixel in the image.

3.1.3 Adaptive Detection

Once we have generated the skin blobs, we use them to build a local statistical model (Line 34, Algorithm 1) that adapts to the particular conditions of the image. From the histogram of these resulting skin regions, we obtain a $P_{local}(c|skin)$. As for non-skin, we assume that the local distribution follows the global one, which gives

$$P_{local}(c|\neg skin) = P_{global}(c|\neg skin) \quad (3.1)$$

The final probability is defined as

$$P(skin|c) = \gamma P_{local}(skin|c) + (1 - \gamma) P_{global}(skin|c) \quad (3.2)$$

where $P_{local}(skin|c)$ and $P_{global}(skin|c)$ are both calculated as in Equation (2.3), differentiating by using local and global data, respectively. The parameter γ controls the importance of the local model.

From Equation (3.2), we generate the final skin probability map, in which the detection can be performed through a fixed threshold or generated by more complex techniques developed for the general probability map. However, the description of such methods is beyond the scope of this paper.

3.2 Experiments

The experiments were evaluated on two different data sets. To train the Bayes classifier, we used 8963 non-skin images and 4666 skin images from the Compaq database [27]. For evaluation and comparison purposes, we used the ECU database [44] divided into a 1000 images for validation and 3000 images for test.

All the following experiments were conducted on an Intel Core i7 3.50GHz with 32GB RAM running 64 bits Ubuntu 12.04 operating system.

In order to determine the bin size of the histogram, we experimented a number of different sizes, as shown in Table 3.1. Since 32 bins per channel produced the highest value of AUC, this value was used in the proposed method both for local and general models.

Table 3.1: Bin size evaluation for validation data set.

Bin size	AUC
8^3	0.892348
16^3	0.918403
32^3	0.934036
64^3	0.923436
128^3	0.917958
256^3	0.914159

Furthermore, the seed detection process demonstrated to be sensitive to the bin size, performing better for more quantized color values. Another factor that influenced the

seeds is the kernel size for the median filter. We empirically observed that the amount of seeds found is directly related to it. If it is too small, then very few seeds are found; otherwise, if it is too large, several false positives are placed as seeds. A 15×15 median filter was used to generate the reported results.

A comparison between our seed extraction method and different fixed thresholds are presented in Table 3.2. Since an important issue in the seed extraction is to avoid false positives while retaining some true positives, the precision (η_{prec}) seems appropriate for the evaluation. As it can be noticed, our method provides superior results with a large difference.

Table 3.2: Evaluation of seed extraction for validation data set.

Method	η_{prec} (%)
$T_{seed} = 0.70$	68.32
$T_{seed} = 0.80$	73.39
$T_{seed} = 0.90$	82.64
$T_{seed} = 0.95$	88.70
Our seeds	94.66

Another desired quality for seeds is that they should be spread over the skin regions to prevent from missing any isolated region. Therefore, we present a qualitative comparison in Figure 3.3. The seeds acquired with fixed thresholds are displayed along with the seeds collected by our method and the T_{seed} found by it are placed in brackets.

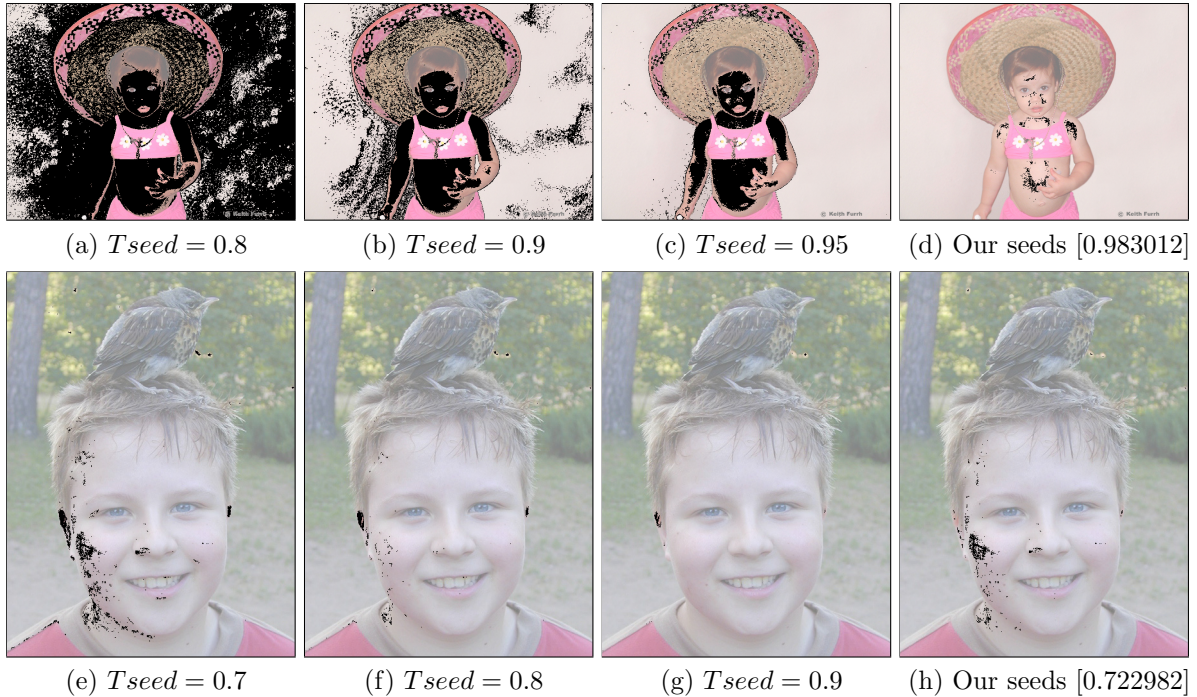


Figure 3.3: Examples of seeds detected through different thresholds for our method.

It is possible to observe that our method not only avoids the non-skin regions but also maintains the dispersal. It is also noticeable that a high threshold is required in the first

image in order to avoid misclassification, while the second will not produce significant seeds for the same value. Thus, our seed extraction method overcomes such problem with a choice of a proper threshold for each image.

The edge detection is an important stage of our method since it supports both seed extraction and propagation. Some experiments were conducted in the edge detector for different color spaces: only luminance, HSV, RGB and YCbCr. The HSV model better captured people's contours in images under abnormal lighting and, therefore, was employed in the experiments. Canny detector was applied with both low and high thresholds equal 100 and a dilation process was performed with a 3×3 structuring kernel.

For comparison, we selected some state-of-the-art methods available in the literature: Cheddad's decision rule [7], statistical model [27], face-based adaptation [29] built with Viola-Jones face-detector [65] and cost propagation [30]. Our method was tested with $\gamma = 0.8$ and $T_{seed_{min}} = 0.5$, whereas original parameters were employed in the other approaches.

Figure 3.4 presents a comparison of the ROC curves. The points in the curves were obtained with different thresholds, except for Cheddad's rule, whose output is binary.

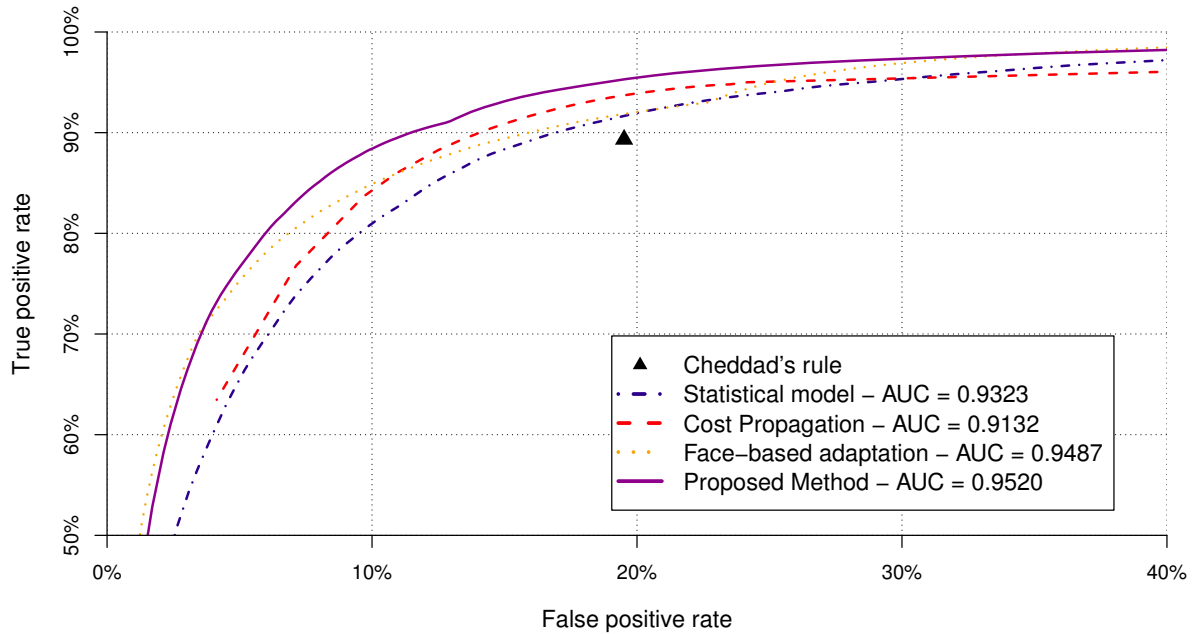


Figure 3.4: ROC curves for comparison of the tested methods.

To present quantitative values, Table 3.3 shows the results for a fixed threshold obtained through maximal Youden's index [72], which represents the closest point to the optimum value (0,1). Best values are highlighted in bold.

Cheddad's rule, as expected, holds the worst results, which demonstrates the inadequacy of such restrict and biased method. Statistical model performs a little better, however, it still has a higher false positive rate.

From the compared methods, the cost propagation has the lowest δ_{min} whereas the face-based adaptation the highest F_{score} . In fact, their values are very similar, considered

Table 3.3: Detection results for different methods.

Method	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)
Cheddad	89.33	19.51	64.78	30.18
Statistical model	87.90	14.51	69.71	26.61
Face-based adaptation	86.83	11.79	72.63	24.96
Cost propagation	90.40	14.46	71.05	24.06
Proposed method	89.78	11.24	74.95	21.46

as a tie. The first has δ_{fp} very similar to the statistical model, what suggests that its improvement is accomplished by an increase in the true positive rate. Oppositely, the second maintains the η_{tp} , performing a reduction in the number of false positives. Our proposed method, which holds the best results, improves on both metrics.

Table 3.4 gives the true positive rate for fixed false positive rate values, where only methods of probabilistic output were considered. It shows the behavior of the methods in different tolerance settings.

Table 3.4: True positive rates for fixed values of false positive rate.

Method	η_{tp} (%)	δ_{fp} (%)	η_{tp} (%)	δ_{fp} (%)
Statistical model	81	10	92	20
Face-based adaptation	85	10	92	20
Cost propagation	84	10	94	20
Proposed method	88	10	95	20

Figure 3.5 illustrates the results for applying the evaluated methods on four different image samples from the test set. In the first row, it is possible to observe that the face-based adaptation misclassified a piece of blue shirt possibly due to the girl’s blue eyes. Furthermore, in the second row shows the detection of several different false positives, which suggests that the face was not correctly detected.

The main drawback of the cost propagation approach is that it detects part of the background as skin, as illustrated in the third row of Figure 3.5. This occurs because, in some points, there is a smooth transition between skin and false skin regions, such that a “leakage” occurs.

Our method overcomes such problems since it tends to use more than just one region for the estimation. Thus, our local model is an accurate representation in cases of variation of skin through different locations. Furthermore, seeds were always found in our tests, while no faces were detected in 12% of the images. Although “leakages” can still occur, they are significantly reduced as demonstrated through the results.

It is also important to highlight the viability of our method in real-time applications, since the average time per image in the test set was 282ms in an unoptimized version of our code.

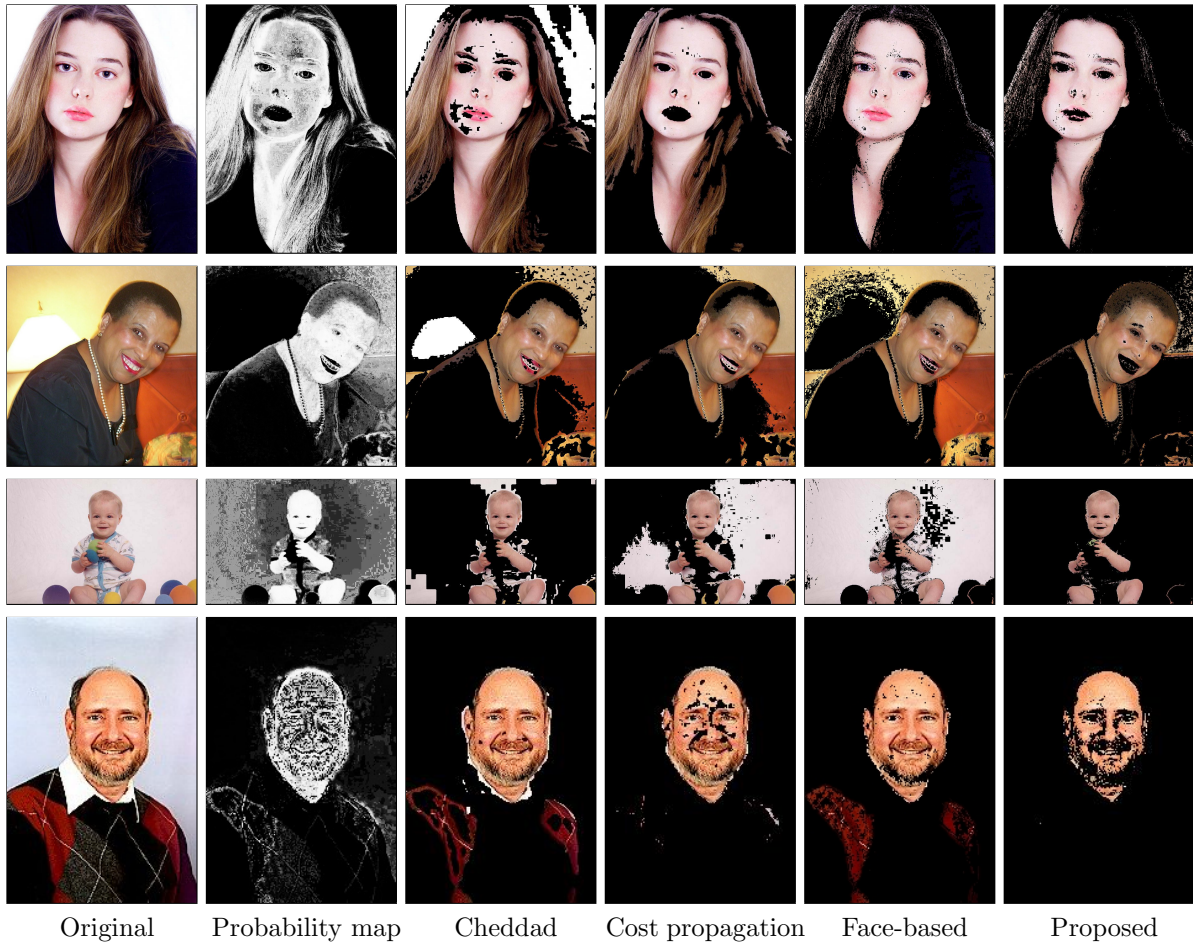


Figure 3.5: Examples of skin detected with different methods.

3.3 Discussion

This work presented a new adaptive human skin segmentation method that makes use of skin probability map and eliminates the need for object detection. The main contributions of our approach include: a new method for seed extraction based on spatial analysis and a self-contained adaptation.

Experimental results demonstrated that the proposed technique outperforms state-of-the-art skin segmentation methods for a large and well-known test set. Nevertheless, additional improvements can be made. False positives generated in the propagation stage of our method have a large contribution to the overall accuracy, what makes us conjecture that even little enhancements in the propagation control, such as better edge detection, will significantly decrease the error rates.

Our method could also be combined with face-based adaptive methods through two strategies: use of detected faces to improve seed extraction and use as an alternative when a face is not present in the image or has not been correctly detected.

As future directions, we intend to investigate more powerful features, such as textural information, to discard incorrect seeds, as well as new strategies for controlling the propagation.

Chapter 4

Human Skin Segmentation Improved by Saliency Detection

In this chapter, we propose the use of saliency to reduce false positives found in probability maps. The objective of the saliency detection process here is to exclude background from the images. Skin probability information is combined with saliency to generate the final skin probability map.

Experiments conducted on large and challenging data sets demonstrate that the proposed method is capable of improving other skin segmentation approaches available in the literature.

The methods developed herein resulted in a paper [50] accepted for presentation in the 16th International Conference on Computer Analysis of Images and Patterns (CAIP'2015).

4.1 Proposed Methodology

We propose a method for reducing the false positive rate in skin segmentation with the use of a saliency detection method. This is based on the premise that the skin is not always salient in the image, but that the background will be not salient. Therefore, saliency detection methods that operate by finding the background to achieve the salient region are preferable, for instance, methods with boundary priors.

Since the skin regions will not be always classified as salient, we need to provide skin information for the considered saliency detector. We deal with skin probability information, however, binary output methods would also be suitable just by considering then having only probability 0 and 1. The main steps of our skin detection framework are illustrated in the diagram of Figure 4.1.

First, the skin detector (Stage 1) is applied to the image, creating a probability map (P_{map}) (Stage 2). This is used to build a weighted image (Stage 3), as shown in Equation 4.1

$$W_I(i, j, k) = P_{map}(i, j) \cdot I(i, j, k) \quad (4.1)$$

where $W_I(i, j, k)$ represents the weighted image pixel in channel k and $I(i, j, k)$ the original image pixel in channel k .

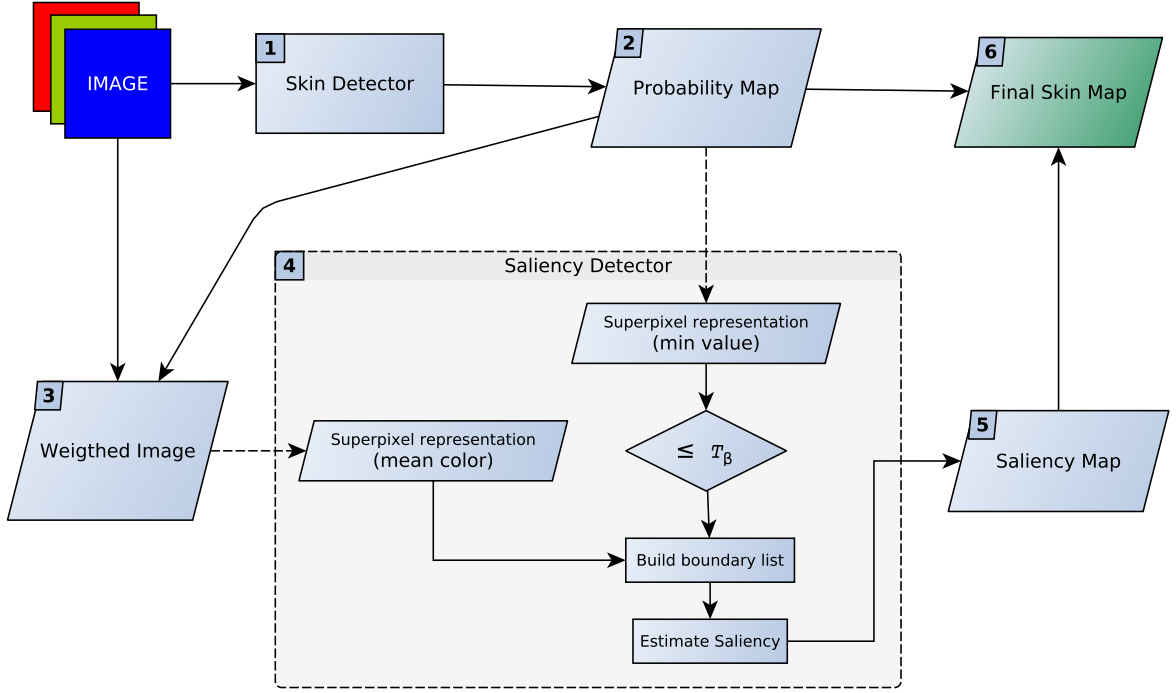


Figure 4.1: Main stages of the proposed skin detection framework.

The weighted image serves as input for the saliency detector (Stage 4), whereas the probability map is also used to exclude probable skin from the boundary list. This is done with a threshold (T_β) applied to the map and aims to prevent that skin pixels adjacent to the boundary will be discarded. Since many saliency implementations use superpixels, in that case the probability map needs to be modeled with the same superpixel structure, however the representative value of each superpixel will be the minimum value of the region instead of the usual mean value. This is done such that only regions containing all probability values larger than T_β will be excluded as background.

The output saliency map (S_{map}) (Stage 5) is again combined with P_{map} , as shown in Equation 4.2

$$F_{map}(i) = \gamma P_{map}(i) + (1 - \gamma) S_{map}(i) \quad (4.2)$$

where F_{map} is the final skin map (Stage 6) and γ defines the weight of the probability map in the mean combination in the range between 0 and 1.

At the final stage, the framework outputs a map for the skin, even for binary skin detectors. Thus, the final segmentation can be performed by a simple threshold or a more sophisticated strategy.

Figure 4.2 illustrates the application of the proposed method to an input image, where the probability map, the weighted image, the saliency map, the final map and the resulting segmentation are shown.

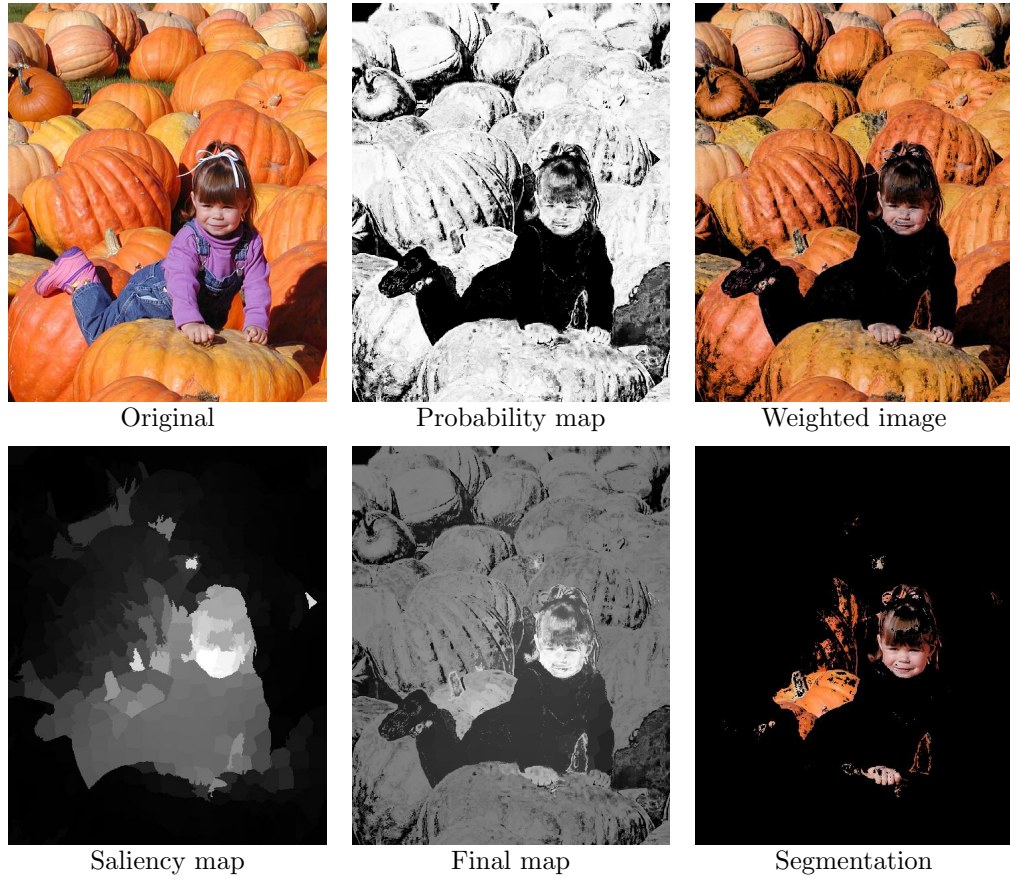


Figure 4.2: Examples of images obtained by applying each stage of our method to an input image.

4.2 Experiments

Experiments were conducted on different data sets to evaluate the proposed methodology. For training, we used 8963 non-skin images and 4666 skin images from the Compaq database [27].

For evaluation and comparison purposes, we used two publicly available skin databases. The ECU database [44] was divided into 1000 images for validation and 3000 images for test, and all the 555 images of IBTD database [74] were used for test.

In order to evaluate the proposed framework, we selected three widely used skin detectors with different approaches: Cheddad’s rule [7] (rule based), Gaussian Mixture Model (GMM) [27] (parametric) and Histogram Model [27] (non-parametric). For the Gaussian Mixture, we used the 16 kernels trained in the original paper for skin and non-skin samples. The Histogram Model was built with 64 bins per channel in the RGB space. It is important to highlight that both GMM and Histogram share the same training set.

As the saliency detector, we chose the saliency optimization method from Zhu et al. [75] since it can cope with the wild scenario of the skin images. In other words, it can work with no centralized foreground and with images of different sizes, preserving the scale and producing less false saliencies. We maintained its original parameters changing only the boundary list as stated in Section 4.1.

Besides the parameters of the chosen skin detectors and saliency detector, our frame-

work adds only two new parameters: the threshold T_β and the weight γ . In the validation stage, we performed a grid search and found that 0.5 is a proper value for both.

Figures 4.3 and 4.4 shows comparative ROC curves between the original skin detector and our combination with saliency detector. It can be seen that the proposed method always achieves superior results on both tested data sets.

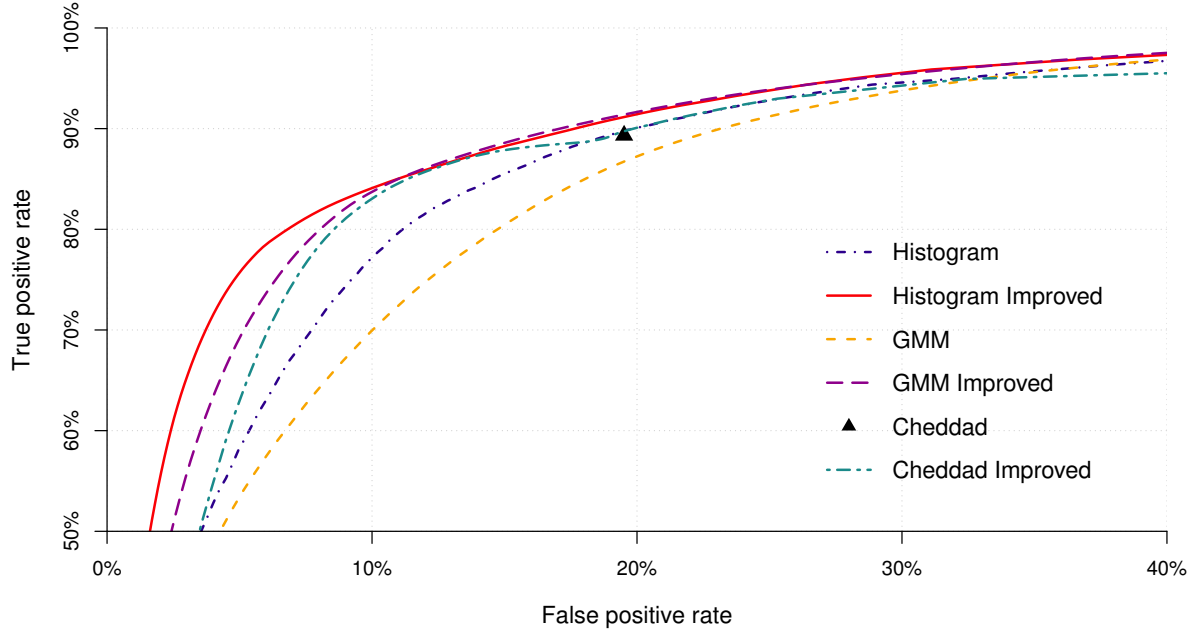


Figure 4.3: ROC curves showing the results on ECU data set for the original method and the improvement with the framework.

Tables 4.1 and 4.2 show the result values when considering the closest point to the optimum point (0, 100%) in the ROC curve. For Cheddad's rule, which is a binary method, the tables present isolated point values. A noticeable aspect in Table 4.1 is that the methods which holds worst results, becomes better than all others without saliency when they are combined by our method with saliency detection.

Table 4.1: Detection results for different methods (ECU data set).

Method	Original				+ Saliency			
	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)
Cheddad	89.33	19.51	64.78	30.18	85.81	12.10	71.67	26.29
Gaussian Mixture	87.55	20.30	63.09	32.76	85.91	11.84	72.08	25.93
Histogram Model	87.21	16.54	66.95	29.33	84.10	10.00	73.63	25.91

For a better view of the differences in performance with and without our proposed methodology, Figure 4.5 shows a graph bar for F-score metric for the original method and its improved version.

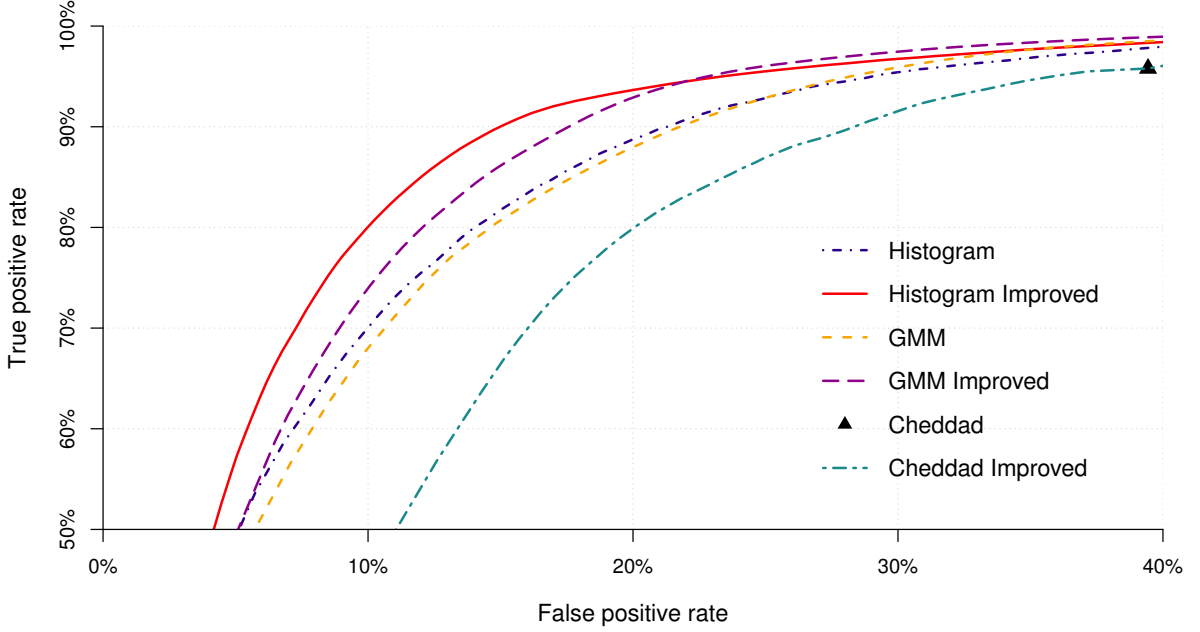


Figure 4.4: ROC curves showing the results on IBTD data set for the original method and the improvement with the framework.

Table 4.2: Detection results for different methods (IBTD data set).

Method	Original				+ Saliency			
	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)
Cheddad	95.74	39.98	53.96	44.24	88.12	26.10	60.36	37.98
Gaussian Mixture	90.83	22.59	64.71	31.76	92.69	19.78	68.30	27.09
Histogram Model	89.70	20.91	65.74	31.21	91.37	16.19	71.44	24.82

For a more detailed comparison, we provide true positive rate values for a 10% false positive rate in Tables 4.3 and 4.4. In other words, this represents how much of true skin is possible to detect since there is only 10% tolerance for skin-like. In case of the original Cheddad method, we perform a linear approximation preserving the same ratio between η_{tp} and δ_{fp} .

It is worth mentioning that our method always results in higher true positive rates with a considerable advantage over the original approaches.

Figure 4.6 shows some examples of final segmentation in the two tested data sets. The first column presents the original image, the second one shows the ground-truth, whereas the remaining columns show the segmentation result for each original method on the left and its correspondent improved segmentation results on the right.

It can be observed that the segmentation results of the methods improved by our framework are clearer and have much less non-skin background. From the last row, it is noticeable that our method not only removes false skin, but also is capable of recovering

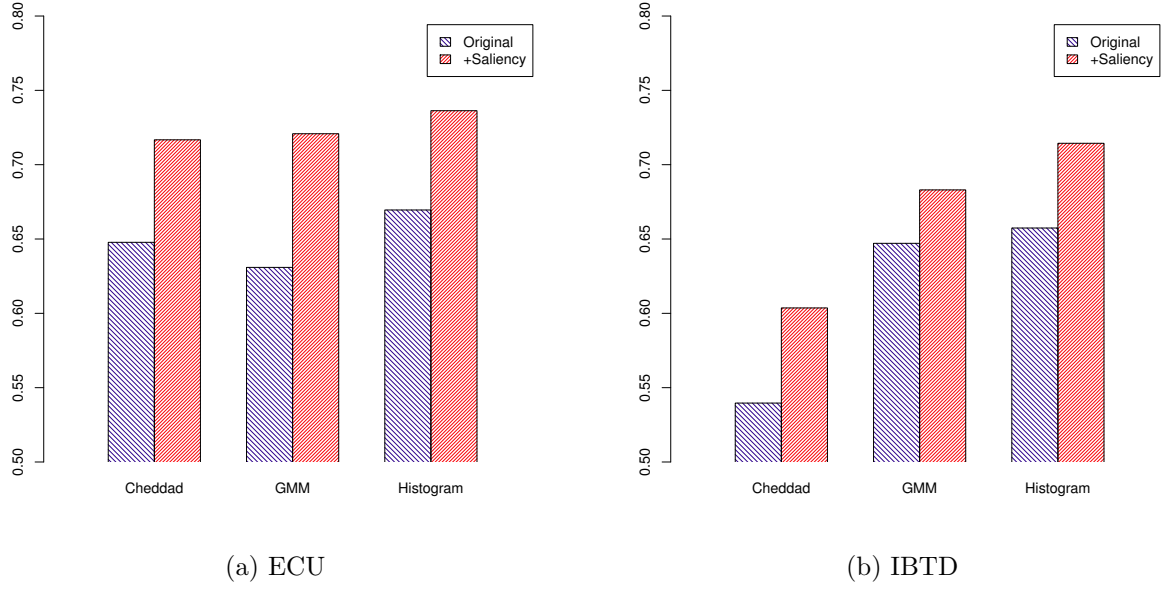


Figure 4.5: Comparative F_{score} results for original methods and our approach both in (a) ECU and (b) IBTD data sets.

Table 4.3: True positive rates for a fixed value of false positive rate (ECU data set).

Method	$\eta_{tp}(\%), \delta_{fp} = 10\%$	
	Original	+ Saliency
Cheddad	46	83
Gaussian Mixture	70	83
Histogram Model	77	84

Table 4.4: True positive rates for a fixed value of false positive rate (IBTD data set).

Method	$\eta_{tp}(\%), \delta_{fp} = 10\%$	
	Original	+ Saliency
Cheddad	24	48
Gaussian Mixture	67	74
Histogram Model	70	80

false negatives.

4.3 Discussion

This work described and analyzed a new general framework for improving skin segmentation using saliency detection. It supports any skin segmentation technique and can be adapted to any saliency detection method based on background priors.

Experiments conducted on two public data sets, a well-known large test set and a more



Figure 4.6: Examples of skin regions detected through different methods.

challenging data set, demonstrated that the proposed technique can provide a significant improvement on the results for different skin detection approaches. Nevertheless, additional refinements can be incorporated into the framework, for instance, in cases where skin covers most of the image, avoiding that true skin is discarded due to its classification as background.

As future directions, we intend to expand the framework to address false positives that inevitably belong to a salient region by removing not only non-skin background, but also non skin present in the foreground.

Chapter 5

Human Skin Segmentation Improved by Texture Energy Under Superpixels

In this chapter, we propose the use of texture energy to reduce the false positives found by color-based methods. Skin color detection is combined with a skin texture probability to generate a final skin probability map.

Experiments conducted on a large and challenging data set demonstrate that the proposed method is capable of improving the skin color segmentation approaches available in the literature.

The methods developed herein resulted in a paper [51] accepted for presentation in the 20th Iberoamerican Congress on Pattern Recognition (CIARP'2015).

5.1 Proposed Methodology

We propose a method for reducing the rate of false positives in skin detection caused by skin-like color. Law's texture energy measure [35] is employed in the process, which works on the response of the intensity image to a special filter mask. The main steps of our skin detection method are illustrated in the flowchart of Figure 5.1.

The filters defined by Law are build by the product of two vectors obtained from a fixed set of 1-D masks designed to detect edges, spots, ripple, among others. A filter is named according the purpose of the vectors from which it was produced and it's size. For example, an E5S5 mask is a 5×5 mask produced by the product of a 1-D edge mask and a 1-D spot mask. In the next section, we explore the choice of the filter for this proposed method.

To allow the calculation of energy over a region and prevent that the same region covers both skin and non-skin, we use the Simple Linear Iterative Clustering (SLIC) [2]. technique for segmenting the image into superpixels. Thus, we calculate the mean energy of each superpixel in the training and test sets.

The goal of the training stage is to obtain two Gaussian models, one for skin and another for non-skin texture energy measures. The images are submitted to superpixels

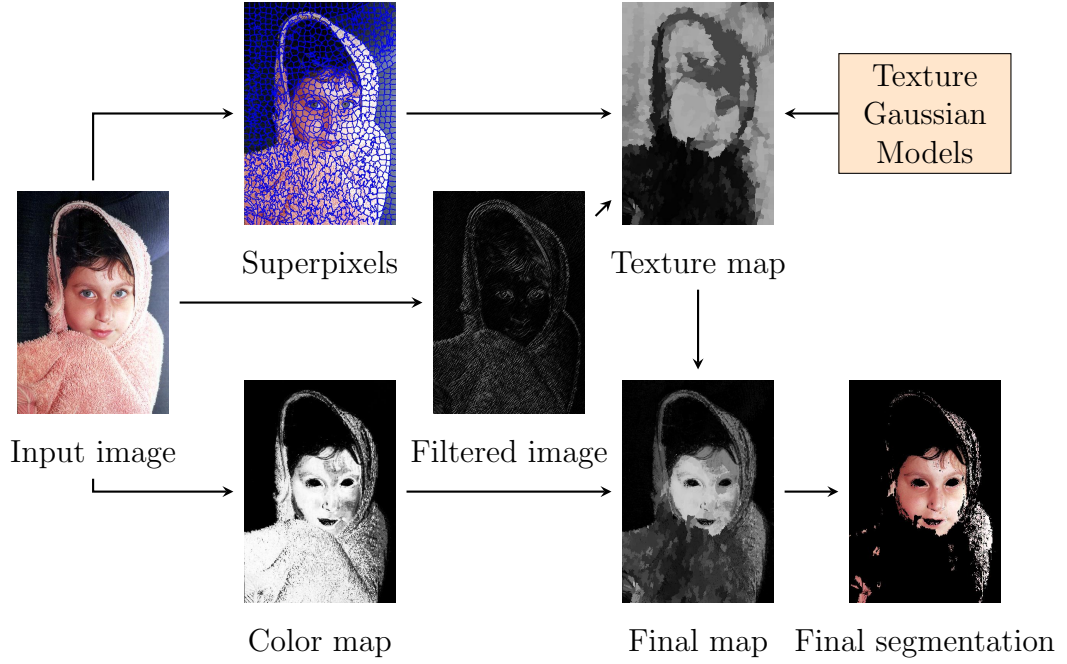


Figure 5.1: Main stages of the proposed skin detection method.

over segmentation and convoluted with a spatial filter. The texture energy is computed for each superpixel, such that mean and standard deviation are extracted for each class (skin and non-skin), forming the two Gaussian models. Algorithm 2 summarizes the training stage.

In the test stage, once the energies of an image have been computed through the same pipeline as in the training step, the skin and non-skin probability densities for each superpixel are obtained. Then, the skin probability given the texture energy is computed in a similar manner to Equation 2.3, as stated in Equation 5.1

$$P(\text{skin}|E_{\Phi}) = \frac{f(E_{\Phi}, \mu_{\text{skin}}, \sigma_{\text{skin}})}{f(E_{\Phi}, \mu_{\text{skin}}, \sigma_{\text{skin}}) + f(E_{\Phi}, \mu_{\neg\text{skin}}, \sigma_{\neg\text{skin}})} \quad (5.1)$$

where E_{Φ} is the energy measure and $f(E_{\Phi}, \mu_{\text{class}}, \sigma_{\text{class}})$ is the Gaussian probability density function for the texture energy.

As texture in a face can vary from the rest of the body, the skin probability in the region close to the nose, around the eyes and mouth will be very low. Thus, it is necessary to apply a heuristic to avoid this type of problem. In our work, we perform a postprocessing mechanism, where areas with low probabilities, surrounded by high probabilities, are filled with the mean of these surroundings high probabilities. Finally, the result of this process constitutes the skin texture probability map.

The texture probability map (T_{map}) is combined with a color probability map (C_{map}) through an AND operation, as shown in Equation 5.2

$$F_{\text{map}} = \sqrt{C_{\text{map}} \cdot T_{\text{map}}} \quad (5.2)$$

producing the final skin probability map F_{map} .

Algorithm 2: Proposed Skin Texture Training.

input : List of images L , list of ground truth pixels G , filter mask f , size of superpixels sp_size .
output: Gaussian Models

```

1  $X_{skin} \leftarrow \emptyset$ 
2  $X_{\neg skin} \leftarrow \emptyset$ 
3 for  $I \in L$  do
4    $SP_{list} \leftarrow SLIC(I, sp\_size)$  /*  $SP_{list}$  holds the superpixels coordinates */
5    $I_{gray} \leftarrow rgb2gray(I)$ 
6    $I_f \leftarrow I_{gray} * f$  /* where  $*$  denotes a convolution */
7   for  $x \in SP_{list}$  do
8      $E_\Phi \leftarrow \frac{\sum I_f(x)^2}{length(SP_{list})}$ 
9     if  $I(x) \in G$  then
10       $X_{skin} \leftarrow \{E_\Phi\}$ 
11    else
12       $X_{\neg skin} \leftarrow \{E_\Phi\}$ 
13    end
14  end
15 end
16 Compute  $\mu$  and  $\sigma$  for  $X_{skin}$  and  $X_{\neg skin}$ .
17 return  $\mu_{skin}, \mu_{\neg skin}, \sigma_{skin}, \sigma_{\neg skin}$ 

```

The color probability map can be calculated from any color skin detector, even binary output methods that produce only probability 0 or 1. At the final stage, the framework outputs a skin map. Thus, the final segmentation can be performed by a simple threshold or a more sophisticated strategy.

5.2 Experiments

Experiments were conducted on two different data sets to evaluate the proposed methodology. For training, we used 8963 non-skin images and 4666 skin images from the Compaq database [27]. For evaluation and comparison purposes, we used the ECU database [44] that was divided into 1000 images for validation and 3000 images for test.

In order to select the filter, we used four 1D vectors:

$$\begin{array}{llll}
\text{L5} & (\text{Level}) & = & \begin{bmatrix} -1 & 4 & 6 & 4 & 1 \end{bmatrix} \\
\text{E5} & (\text{Edge}) & = & \begin{bmatrix} -1 & -2 & 0 & 2 & 1 \end{bmatrix} \\
\text{S5} & (\text{Spot}) & = & \begin{bmatrix} -1 & 0 & 2 & 0 & 1 \end{bmatrix} \\
\text{R5} & (\text{Ripple}) & = & \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \end{bmatrix}
\end{array}$$

which generates sixteen 5×5 filters. Each one is convoluted with the image; the results of filters that are just transposed of others are combined to produce their mean, resulting in nine features. From the validation set, we evaluate the gain provided by each feature in relation to the color individually. The filter E5S5/S5E5 produced the best results.

$$E5^t \cdot S5 = \begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 2 & 0 & -4 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 4 & 0 & -2 \\ -1 & 0 & 2 & 0 & -1 \end{bmatrix}$$

In order to evaluate the proposed method, we selected three widely used skin detectors with different approaches: Cheddad's rule [7] (rule based), Histogram Model [27] (non-parametric) and Gaussian Mixture Model (GMM) [27] (parametric). The Histogram Model was built with 64 bins per channel in the RGB space. For the Gaussian Mixture, we used the 16 kernels trained in the original paper with the same database as used here.

Figure 5.2 shows comparative ROC curves between the original skin detector and our improvement. It is possible to observe that the proposed method always achieves superior results.

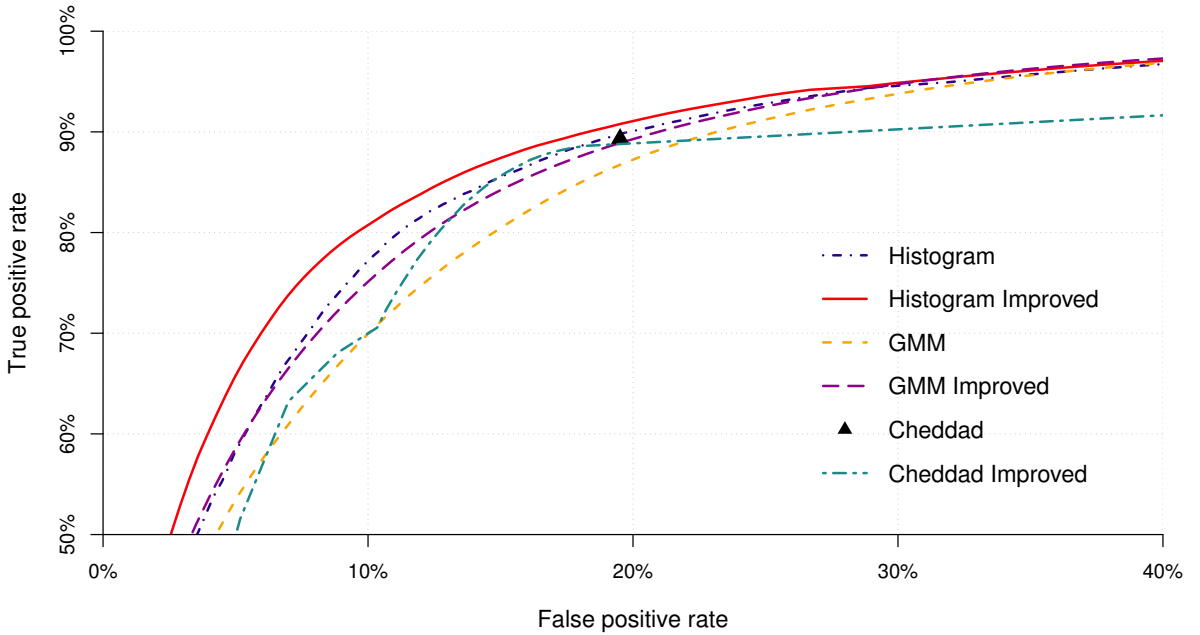


Figure 5.2: ROC curve illustrating the results on test data set for the original method and the improvement through our method.

Table 5.1 shows the result values when considering the closest point to the optimum point (0, 100%) in the ROC curve. For Cheddad's rule, which is a binary method, the tables present isolated point values.

For a more detailed comparison, we provide true positive rate values for a 10% false positive rate in Table 5.2. In other words, this represents how much of true skin is possible to detect since there is only 10% tolerance for skin-like. In case of the original Cheddad method, we perform a linear approximation preserving the same ratio between η_{tp} and δ_{fp} .

It is worth mentioning that our method always results in higher true positive rates

Table 5.1: Detection results for different methods (ECU data set).

Method	Original				Improved			
	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)	η_{tp} (%)	δ_{fp} (%)	F_{score} (%)	δ_{min} (%)
Cheddad	89.33	19.51	64.78	30.18	87.32	16.22	67.38	28.90
Gaussian Mixture	87.55	20.30	63.09	32.76	87.37	17.78	65.64	30.41
Histogram Model	87.21	16.54	66.95	29.33	86.96	14.55	69.17	27.59

with a considerable advantage over the original approaches.

Table 5.2: True positive rates for a fixed value of false positive rate (ECU data set).

Method	$\eta_{tp}(\%), \delta_{fp} = 10\%$	
	Original	Improved
Cheddad	46	71
Gaussian Mixture	70	75
Histogram Model	77	81

Figure 5.3 shows some examples of final segmentation in the tested data set. The first column presents the original image, the second one shows the ground-truth, whereas the remaining columns show the segmentation result for each original method on the left and its correspondent improved segmentation results on the right.



Figure 5.3: Examples of skin regions detected through different methods.

5.3 Discussion

This work described a new technique for reducing the high number of false positives in color-based skin segmentation. It can be applied in conjunction with any skin segmentation method, while not adding more sensitive parameters to perform the skin classification.

Experiments conducted on a well-known large test set demonstrated that the proposed technique can provide a significant improvement on the results obtained with different color-based skin detection approaches. Furthermore, the use of texture for skin segmentation is very difficult; although human skin has a distinguishable texture, this can only be noticeable on high resolution images. Age, amount of body hair, expression and pose constitute obstacles in finding a structural pattern for skin.

As directions for future work, we intend to expand the method with multi-resolution analysis to take account into the variety in quality and size of the images. We also plan to propose a better filter, designed specifically for the skin detection problem.

Chapter 6

Conclusions and Future Work

In this work, a set of improvements were proposed to satisfy the need of increasing the separability between skin and skin-like regions for human skin segmentation. Three main contributions are given: a self-adaptive skin color model, a method that use saliency for non-skin background removal and a combination of skin texture and color.

The self-adaptive proposed method fits a skin color model to particular conditions of the images, addressing the problem of lighting variation and natural differences that occur in skin colors among people. Although this approach is not novel, our method avoids the use of weak heuristics and relies only on the skin color itself. Experiments demonstrated that it overcomes not only non-adaptive methods but also an adaptive method based on faces. This is explained due to the inevitable errors brought by the face detector and because faces contain more than just skin.

The improvement by using the concept of saliency detection is a novelty since it has not been well explored in the skin segmentation problem. The use of saliency has many advantages: no training is necessary, it is fast and there are many different methods that can be employed. It was possible to observe from the experiments that the use of saliency always improved the results of color-based detectors. The weaker the detector is, the larger is the improvement.

The use of texture for skin detection is a complex task. The image quality, person's pose, age and amount of hair represent a major obstacle to define a skin pattern. Thus, our method focuses on simple features, such as coarseness and homogeneity. Nonetheless, they are not explicitly defined but learned from a training set. In the end, it is observed that texture detectors work as an anomaly detection process, eliminating non-skin regions that are very different from the skin. The use of superpixels makes the process more accurate since each evaluated region must contain only skin or non-skin pixels. If a region presents both skin and non skin portions, any classification method will fail. Experimental results showed that any type of approach using color can be combined with our texture classification, providing an overall improvement.

Each of these three methods have theirs advantages and drawbacks in relation to the others. The self-adaptive is not suitable when the skin color varies a lot in the same image, but is excellent for data sets with high intra-image skin color variation. The method with saliency is useful for images where the skin will be centralized and most the image constitutes of non-skin background. The texture energy method is the one with

the low results in comparison with the first two, this happens because it operates in a more specific space. So the accuracy depends on the textural difference between skin and non skin in the data sets. It will be more suitable for images where the background is not uniform, like natural images.

As directions for future work, we intend to propose a skin detection framework by combining all the multiple evidences presented here. It is expected that the combination of the methods will probably overcome the obtained individual results. For instance, saliency could be used to limit the propagation step of the self-adaptive method. Furthermore, texture energy could be used to better select seeds.

Finally, we plan to apply and evaluate our approaches to other problems, such as gesture detection, nudity identification, face detection, among others.

Bibliography

- [1] R. Achanta, F. Estrada, P. Wils, and S. Ssstrunk. Salient Region Detection and Segmentation. In *Computer Vision Systems*, pages 66–75. Springer, 2008.
- [2] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [3] J. Ahlberg. A System for Face Localization and Facial Feature Extraction. Technical Report LITH-ISY-R-2172, Department of Electrical Engineering, Linkping University, Linkping, Sweden, 1999.
- [4] S. Bilal, R. Akmeliawati, M. J. E. Salami, and A. A. Shafie. Dynamic Approach for Real-Time Skin Detection. *Journal of Real-Time Image Processing*, pages 1–15, 2012.
- [5] L. Bretzner, I. Laptev, and T. Lindeberg. Hand Gesture Recognition using Multi-scale Colour Features, Hierarchical Models and Particle Filtering. In *Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 423–428. IEEE, 2002.
- [6] D. Chai and K. N. Ngan. Face Segmentation using Skin-Color Map in Videophone Applications. *IEEE Transactions on Circuits and Systems for Video Technology*, 9(4):551–564, 1999.
- [7] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt. A Skin Tone Detection Algorithm for an Adaptive Approach to Steganography. *Signal Processing*, 89(12):2465–2478, 2009.
- [8] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu. Global Contrast Based Salient Region Detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 409–416. IEEE, 2011.
- [9] R. S. Chora. CBIR system for detecting and blocking adult images. In *9th WSEAS International Conference on Signal Processing*, pages 52–57, Stevens Point, Wisconsin, USA, 2010. World Scientific and Engineering Academy and Society (WSEAS).
- [10] D. Comaniciu and P. Meer. Mean shift: A Robust Approach Toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

- [11] T. Darrell, G. Gordon, M. Harville, and J. Woodfill. Integrated Person Tracking using Stereo, Color, and Pattern Detection. *International Journal of Computer Vision*, 37(2):175–185, 2000.
- [12] E. W. Dijkstra. A Note on Two Problems in Connection with Graphs. *Numerische Mathematik*, 1:269–271, 1959.
- [13] P. Dollar, Z. Tu, and S. Belongie. Supervised Learning of Edges and Object Boundaries. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 1964–1971. IEEE, 2006.
- [14] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Graph-based Image Segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [15] J. Fritsch, S. Lang, M. Kleinhagenbrock, G. A. Fink, and G. Sagerer. Improving Adaptive Skin Color Segmentation by Incorporating Results from Face Detection. In *11th IEEE International Workshop on Robot and Human Interactive Communication*, pages 337–343, 2002.
- [16] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-Aware Saliency Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10):1915–1926, 2012.
- [17] R. Gonzalez, R. Woods, and S. Eddins. *Digital Image Processing using MATLAB*. Gatesmark Publishing, 2009.
- [18] C. Guo, Q. Ma, and L. Zhang. Spatio-Temporal Saliency Detection using Phase Spectrum of Quaternion Fourier Transform. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
- [19] T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The Elements of Statistical Learning*, volume 2. Springer, 2009.
- [20] X. Hou, J. Harel, and C. Koch. Image Signature: Highlighting Sparse Salient Regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):194–201, 2012.
- [21] X. Hou and L. Zhang. Saliency Detection: A Spectral Residual Approach. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [22] S. Hsieh, K.-C. Fan, and C. Lin. A Statistic Approach to the Detection of Human Faces in Color Nature Scene. *Pattern Recognition*, 35(7):1583–1596, 2002.
- [23] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):696–706, 2002.
- [24] X. Hu, S. Peng, J. Yan, and N. Zhang. Fast Face Detection based on Skin Color Segmentation using Single Chrominance Cr. In *7th International Congress on Image and Signal Processing*, pages 687–692. IEEE, 2014.

- [25] S. Ji, X. Lu, and Q. Xu. A Fast Face Detection Method Combining Skin Color Feature and AdaBoost. In *International Conference on Multisensor Fusion and Information Integration for Intelligent Systems*, pages 1–5. IEEE, 2014.
- [26] Z. Jiang, M. Yao, and W. Jiang. Skin Detection using Color, Texture and Space Information. In *Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, volume 3, pages 366–370, 2007.
- [27] M. J. Jones and J. M. Rehg. Statistical Color Models with Application to Skin Detection. *International Journal of Computer Vision*, 46(1):81–96, 2002.
- [28] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A Survey of Skin-Color Modeling and Detection Methods. *Pattern Recognition*, 40(3):1106–1122, 2007.
- [29] M. Kawulok. Energy-based Blob Analysis for Improving Precision of Skin Segmentation. *Multimedia Tools and Applications*, 49(3):463–481, 2010.
- [30] M. Kawulok. Fast Propagation-based Skin Regions Segmentation in Color Images. In *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pages 1–7, 2013.
- [31] M. Kawulok, J. Nalepa, and J. Kawulok. Skin Detection and Segmentation in Color Images. In *Advances in Low-Level Color Image Processing*, pages 329–366. Springer, 2014.
- [32] J. Kovac, P. Peer, and F. Solina. Human Skin Color Clustering for Face Detection. In *International Conference on Computer as a Tool*, volume 2, pages 144–148. IEEE, Sept. 2003.
- [33] H. Kruppa, M. A. Bauer, and B. Schiele. Skin Patch Detection in Real-World Images. In *Pattern Recognition*, pages 109–116. Springer, 2002.
- [34] G. Kukharev and A. Nowosielski. Fast and Efficient Algorithm for Face Detection in Colour Images. *Machine Graphics & Vision International Journal*, 13(4):377–399, 2004.
- [35] K. I. Laws. Rapid Texture Identification. In *24th Annual Technical Symposium*, pages 376–381. International Society for Optics and Photonics, 1980.
- [36] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen. Naked Image Detection based on Adaptive and Extensible Skin Color Model. *Pattern Recognition*, 40(8):2261–2270, 2007.
- [37] J.-Y. Lee and S. I. Yoo. An Elliptical Boundary model for Skin Color Detection. In *International Conference on Imaging Science, Systems, and Technology*. Citeseer, 2002.
- [38] A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi. Turbopixels: Fast Superpixels using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2290–2297, 2009.

- [39] S. J. McKenna, S. Gong, and Y. Raja. Modelling Facial Colour and Identity with Gaussian Mixtures. *Pattern Recognition*, 31(12):1883–1892, 1998.
- [40] A. P. Moore, S. Prince, J. Warrell, U. Mohammed, and G. Jones. Superpixel Lattices. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
- [41] G. Mori, X. Ren, A. A. Efros, and J. Malik. Recovering Human Body Configurations: Combining Segmentation and Recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages II–326. IEEE, 2004.
- [42] P. Ng and C.-M. Pun. Skin Color Segmentation by Texture Feature Extraction and K-means Clustering. In *Third International Conference on Computational Intelligence, Communication Systems and Networks*, pages 213–218. IEEE, 2011.
- [43] H. Pedrini and W. Schwartz. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Editora Thomson Learning, 2007.
- [44] S. L. Phung, A. Bouzerdoun, and D. Chai Sr. Skin Segmentation using Color Pixel Classification: Analysis and Comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):148–154, 2005.
- [45] S. L. Phung, D. Chai, and A. Bouzerdoun. Adaptive Skin Segmentation in Color Images. In *International Conference on Multimedia and Expo*, volume 3, pages III–173, 2003.
- [46] C. Platzer, M. Stuetz, and M. Lindorfer. Skin Sheriff: A Machine Learning Solution for Detecting Explicit Images. In *2nd International Workshop on Security and Forensics in Communication Systems*, pages 45–56, New York, NY, USA, 2014. ACM.
- [47] X. Ren and J. Malik. Learning a Classification Model for Segmentation. In *Ninth IEEE International Conference on Computer Vision*, pages 10–17. IEEE, 2003.
- [48] J. Ruiz-del Solar and R. Verschae. Skin Detection using Neighborhood Information. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 463–468. IEEE, 2004.
- [49] A. Santos and H. Pedrini. A Self-Adaptation Method for Human Skin Segmentation Based on Seed Growing. In *10th International Conference on Computer Vision Theory and Applications*, pages 455–462, Berlin, Germany, Mar. 2015.
- [50] A. Santos and H. Pedrini. Human Skin Segmentation Improved by Saliency Detection. In *Computer Analysis of Images and Patterns*, pages 1–12, Valetta, Malta, Sept. 2015.
- [51] A. Santos and H. Pedrini. Human Skin Segmentation Improved by Texture Energy under Superpixels. In *20th Iberoamerican Congress on Pattern Recognition*, Montevideo, Uruguay, Nov. 2015.

- [52] F. Saxen and A. Al-Hamadi. Color-Based Skin Segmentation: An Evaluation of the State of the Art. In *IEEE International Conference on Image Processing*, pages 4467–4471. IEEE, 2014.
- [53] H. Sevimli, E. Esen, T. K. Ateş, E. C. Ozan, M. Tekin, K. B. Loğoğlu, A. M. Sevinç, A. Saracoğlu, A. Yazici, and A. A. Alatan. Adult Image Content Classification using Global Features and Skin Region Detection. In *Computer and Information Sciences*, pages 253–258. Springer, 2010.
- [54] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.
- [55] P. M. Silva and M. C. Polastro. An Overview of NuDetective Forensic Tool and its Usage to Combat Child Pornography in Brazil. In *XLIII Jornadas Argentinas de Informática e Investigación Operativa - VI Workshop de Seguridad Informática*, 2014.
- [56] K. Sobottka and I. Pitas. Face Localization and Facial Feature Extraction based on Shape and Color Information. In *International Conference on Image Processing.*, volume 3, pages 483–486. IEEE, 1996.
- [57] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen. Skin Detection in Video under Changing Illumination Conditions. In *15th International Conference on Pattern Recognition*, volume 1, pages 839–842. IEEE, 2000.
- [58] H. Stern and B. Efron. Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 249–254. IEEE, 2002.
- [59] R. Subban and R. Mishra. Human Skin Segmentation in Color Images using Gaussian Color Model. In *Recent Advances in Intelligent Informatics*, pages 13–21. Springer, 2014.
- [60] M. J. Taylor and T. Morris. Adaptive Skin Segmentation via Feature-based Face Detection. In *SPIE Photonics Europe*, page 91390P. International Society for Optics and Photonics, May 2014.
- [61] J.-C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu. Comparative Performance of Different Skin Chrominance Models and Chrominance Spaces for the Automatic Detection of Human Faces in Color Images. In *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 54–61. IEEE, 2000.
- [62] S. Tsekeridou and I. Pitas. Facial Feature Extraction in Frontal Views using Biometric Analogies. In *European Signal Processing Conference*, volume 98, pages 315–318, 1998.
- [63] V. Vezhnevets, V. Sazonov, and A. Andreeva. A Survey on Pixel-based Skin Color Detection Techniques. In *Graphicon*, volume 3, pages 85–92. Moscow, Russia, 2003.

- [64] L. Vincent and P. Soille. Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.
- [65] P. Viola and M. J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [66] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based Hand-Gesture Applications. *Communications of the ACM*, 54(2):60–71, Feb. 2011.
- [67] X. Wang, X. Zhang, and J. Yao. Skin Color Detection under Complex Background. In *International Conference on Mechatronic Science, Electric Engineering and Computer*, pages 1985–1988, 2011.
- [68] Y. Wang and B. Yuan. A Novel Approach for Human Face Detection from Color Images under Complex Background. *Pattern Recognition*, 34(10):1983–1992, 2001.
- [69] Y. Wei, F. Wen, W. Zhu, and J. Sun. Geodesic Saliency using Background Priors. In *European Conference on Computer Vision*, pages 29–42. Springer, 2012.
- [70] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang. Saliency Detection via Graph-based Manifold Ranking. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3166–3173. IEEE, 2013.
- [71] M.-H. Yang and N. Ahuja. Gaussian Mixture Model for Human Skin Color and its Application in Image and Video Databases. In *SPIE: Storage and Retrieval for Image and Video Databases VII*, volume 3656, pages 458–466, 1999.
- [72] W. J. Youden. Index for Rating Diagnostic Tests. *Cancer*, 3(1):32–35, 1950.
- [73] M.-J. Zhang and W. Gao. An Adaptive Skin Color Detection Algorithm with Confusing Backgrounds Elimination. In *IEEE International Conference on Image Processing.*, volume 2, pages II–390. IEEE, 2005.
- [74] Q. Zhu, C.-T. Wu, K.-T. Cheng, and Y.-L. Wu. An Adaptive Skin Model and its Application to Objectionable Image Filtering. In *12th annual ACM International Conference on Multimedia*, pages 56–63. ACM, 2004.
- [75] W. Zhu, S. Liang, Y. Wei, and J. Sun. Saliency Optimization from Robust Background Detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2814–2821. IEEE, 2014.