FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Automatic human-head and shoulder segmentation of frontal-view face images

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Visual Computing

eingereicht von

Robin Melán, B.Sc.
Matrikelnummer 1029201

an der Fakultät für Informatik
der Technischen Universität Wien
Betreuung: O. Univ. Prof. Dr. Walter Kropatsch

Wien, 1. Jänner 2018

Robin Melán	Walter Kropatsch

FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Automatic human-head and shoulder segmentation of frontal-view face images

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Visual Computing

by

Robin Melán, B.Sc.
Registration Number 1029201

to the Faculty of Informatics

at the TU Wien

Advisor: O. Univ. Prof. Dr. Walter Kropatsch

Vienna, 1st January, 2018

Robin Melán

Walter Kropatsch

Erklärung zur Verfassung der Arbeit

Robin Melán, B.Sc.
Oberzellergasse 3/17/12, 1030 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

Wien, 1. Jänner 2018

Robin Melán

Kurzfassung

Bildsegmentierung ist eines der elementaren Themen in der Mustererkennung und Bildverarbeitung. Das erst kürzlich entdeckte spezifische Problem einer automatischen Kopf-, Gesicht- und Schultersegmentierung, gewinnt in letzter Zeit an Bedeutung für eine Reihe von Anwendungen. Vorranging wäre eine automatische Extraktion einer Person von einem undefinerten, komplexen Hintergrund für jegliche Profilbilder von Nutzen, die in Dokumenten verwendet werden können. Ebenfalls würde eine solche Anwendung Verbesserungen bei der automatischen Gesichtserkennung bedeuten, und weiters im E-Government Bereich bzw. gewerblichen Sektor von Interesse sein. In dieser Arbeit behandeln wir das Problem der automatischen Kopf-, Gesicht- und Schultersegmentierung aus Bildern in Frontalansicht mit einem undefinierten komplexen Hintergrund, indem wir eine Methodik bestehend aus individuellen Teilaufgaben präsentieren. Diese Teilaufgaben können unabhängig voneinander betrachtet werden und bestehen aus einer Gesichtshautfarbe-Detektion, einer Gegenüberstellung zweier Superpixel Algorithmen und der Untersuchung von Haar- und Kleidungseigenschaften für unsere Haar- und Schultersegmentierung. Wir evaluieren unsere Methoden und präsentieren konkurrenzfähige Resultate in jeder Teilaufgabe.

Abstract

Object segmentation is one of the basic issues in image processing and computer vision. However, especially human-head and shoulder segmentation is a topic which was introduced only recently, gaining in importance for a wide area of computer vision applications, such as testing compliance for ID document issuing, improving images for facial recognition or even used in the upcoming e-governmental self services and commercial sector. In this thesis we address the problem of automatic human-head and shoulder segmentation of frontal-view face images from non-uniform complex backgrounds and propose an approach composed of different subtask. These subtasks can be viewed individually and consist of a novel face skin silhouette detection approach based on supervised classification learners, a study of two state-of-the-art superpixel algorithms in relation to the specified problem statement and a novel hair and shoulder segmentation approach. We discuss and evaluate our methods and present competitive results for each subtask.

Contents

CHAPTER 1

Introduction

In this thesis, we focus on a particular segmentation task: extracting the human-head and shoulder boundary of static frontal-view face images from an arbitrary complex background.
The automatic detection and segmentation of human subjects for static images is still a challenge, due to several real world factors such as illumination conditions, shadows, occlusions and background clutter, unnatural skin tones, different ethnicity groups, color saturation and no prior knowledge about the person nor its environment and background in the image. Another challenge to be mentioned are problems like image quality, image noise, resolution, or even issues related to factors associated with the dynamic of the human being, such as the great variety of poses, appearance and shapes [20].
Automatic people detection and segmentation in general can be widely used in many computer vision based applications, including photo analysis, surveillance systems, people counting, robotics, natural user interfaces and editing [21]. The particular problem of human-head and shoulder segmentation especially can be of interest for a wide range of applications.

1.1 Motivation

As previously mentioned, the idea of having such a segmented portrait image of a persons head and shoulders with a uniform background can be of interest for many different kinds of applications and areas.

The ISO [1]/IEC 19794-5 Information technology — Biometric data interchange formats — Part 5: Face image data [19], is the fifth of 8 parts of the ISO standard ISO/IEC 19794, published in 2005 by the International Civil Aviation Organization (ICAO). It describes interchange formats for several types of biometric data, more specifically defining a

[1] International Organization for Standardization

standard scheme for codifying data describing human faces to be used correctly by facial recognition systems. Modern biometric passport photos should comply with this standard [18].
Many organizations and public authorities have already started enforcing its directives, and several software applications have been produced, like the ICAO Portrait Checker Module [2], to automatically test compliance to the specification. These software applications provide automatic checks on the head pose, head position, occlusion, expression, eye visibility, illumination artifacts caused by glasses, illumination and color condition, specifically color saturation (over or under exposure) and unequal light distribution on the face. The examination whether the ICAO requirement for a uniform background is met has still to be done **manually** by an official. So to issue a passport or any other kind of document e.g. driving license, credit card, personal identity card, etc. public authorities or cooperations inspect the submitted profile picture initially manually for background uniformity and then check the other requirements with an ICAO software to verify if all other criteria are given [2]. The uniform background is important for further computer face recognition algorithms, because the first step is to compute facial features by extracting landmarks from the subject's face to then compare them with a face database. The uniform background improves the identification rate in the face recognition process, since the confusion of false determination of facial landmarks in the background is reduced or even removed.
Another problem occurs for countries and their public authorities or cooperations which do not follow the ISO/IEC 19794 norm yet. In the past they accepted any kind of images not following this criteria using face images with random background or even scanned pictures. The ability to reuse these images as references for an automatic face recognition software requires human-head and shoulder segmentation as well.

Other possible cases where such a solution is needed lie in the upcoming e-government sector. A filing of application, like a passport, personal identity or any kind of document renewal could be initiated online, uploading a digital photo (e.g. taken with a webcam) which is then checked by an ICAO-Checker and beforehand segmented with the proposed segmentation algorithm from this master thesis.
Similarly to the idea of usage for e-government self service, a software including the proposed methodology could be created for the event management area, issuing out identification badges (e.g. VIP, guest pass, host badge) which contain a photograph for the registration process. To facilitate the manual verification the background would be removed.

In future work this idea could be extended for applications in the law enforcement sector, identifying people which are on a blacklist via surveillance cameras, a 1:N verification. The video stream could be split into frames, choosing the one with optimal face detection. The further segregation of the image background would increase the success rate of face recognition algorithms, reducing the considerably large hit list due to low resolution and

[2] Manual Biometrics: ICAO Portrait Checker technical description from the Biometrics Center Atos, June 2011

quality.

Similarly, this idea could be used in scope of commercial applications using white lists,e.g. in shopping centers, recognizing noted customers as they enter and notifying them about offers, news, etc.

1.2 Problem Statement

Recording a still portrait image can be very diverse and variable. Considering the problem from a computer vision perspective, finding an accurate head and shoulder silhouette in static images is difficult due to several real world factors such as different illumination conditions, which include shadows, unnatural skin tones, saturated colors (e.g. too dark, too light), as well as occlusion, background clutter, image quality, image noise, low resolution, variety of head poses, appearance, shape and the randomly textured background.

The aim of this master thesis is to study the problem of automatic human-head and shoulder segmentation of frontal-view face images and analyze the proposed methodological strategy. We provide an approach which detects and segments out the arbitrary background to be replaced with a uniform background, to achieve the expected result showed in Figure 1.1. For this we allow to define some technical specifications and

(1) (2)

Figure 1.1: (1) Input image. (2) Ground truth output image.

conditions for the input image, which will play a role in the individual topics studied in the main section of this thesis. Some of these conditions follow the defined data requirements stated in the ISO/IEC 19794-5 standard [18].

- *Pose:* Rotation of the head shall be less than $+/-5$ degrees from frontal in every direction roll, pitch and yaw. Specially when it comes to face recognition software, the pose is known to strongly affect performance of automated face recognition systems.

- *Occlusion:* As we define our interest in frontal-view face images where the person is looking towards the camera, we exclude any sort of occlusion of the persons face, specially facial landmarks like eyes, nose and mouth.

- *Eye glasses lightning artifacts:* If the person wears glasses, they shall be clear glass and transparent so that the eyes are clearly visible. Furthermore no lightning artifacts or flash reflections on glasses shall be projected on to the image.

- *Over or under exposure:* The gradations in textures for the persons face, hair and clothes shall be clearly visible. In this sense, the pictures will be within a range of saturation without over or under exposure. On the one hand, if the exposure is too long or lens aperture opened too widely, then the image is over exposed and too bright. The the other hand, if exposure is too short or lens aperture too small, then the image is under exposed and therefore too dark.

- *Focus and depth of field:* The subject's captured image shall always be in focus from nose to ears and chin to crown.

- *Unnatural color:* The illumination shall produce a face image with natural looking flesh tones when viewed in typical examination environments.

- *Resolution:* The resolution for a frontal-view face image, which consists of face, hair and outline of the shoulders of the subject, must be 420x525 pixels as minimum. Moreover, the eye distance has to be at least 90 pixels. These requirements ensure a certain face and hair quality to be able to extract meaningful information, but also assures the possibility to use the output portrait results as profile pictures in documents according to the ISO/IEC 19794-5 standard [18] if desired. Additional artificial changes of color representation, contrast, focus and intensity, such as changes e.g. for purpose of beauty enhancement are not included.

- *No baldness:* For future work we leave open the possibility to handle bald and semi-bald people, so in our proposed method we are considering strictly people with hair.

- *Aesthetic look:* The goal is to achieve an output result which allows a correct face and shoulder segmentation with a detailed and aesthetic representation of the hair. Since these results could land on documents, which often are checked for identification manually by an officer, the aesthetic plays an important role as well. So in the context of this thesis an aesthetic result of the hair and shoulder structure, means the segmented hair and outline of the shoulders, has to look *normal* in the eyes of the beholder and, for instance, it is not important for every single strand of hair to be segmented correctly.

- *Background complexity:* Our interest lies in handling a human-head and shoulder segmentation where the background is not uniform, but complex, including indoor/outdoor scenes and shadows, whilst a contrast from the persons face, hair

and clothes in relation to the surrounding background is given (e.g. black hair on black background is excluded in this thesis).

- *Biometric features:* Important to mention is that all the biometric features of the person remain unchanged. There are no modifications done on the persons properties in the image.

1.3 Contribution

Our contributions in this thesis are the following:

- Proposing a methodology for human-head and shoulder segmentation of frontal view face images, by splitting up the complex problem into different subtasks and proposing an approach to solve each subtask. These approaches combined give a possible methodology to this particular problem statement, but can be observed individually as well and applied independently for other tasks with similar technical requirements.
- Introducing a novel skin detection approach based on classification learners, extending the training set of the classifiers by adding automatically labeled subset of pixels extracted from the test image.
- Comparing two very different superpixel algorithms which oversample the image, with one adhering the topology and the other resulting into compact similar sized homogeneous regions to simplify the extraction of features.
- Characterizing hair, clothes and the background in an oversampled image with color, texture and the superpixels relative position generating hair, shoulder and background models without any prior knowledge on the person and background complexity of the image.
- Labeling due to occlusion of not connected hair regions as one class.
- Maintain the biometric features of the person unchanged to allow an identity check in the following.

1.4 Structure of the Thesis

The remainder of this report is organized as follows: Chapter 2 gives a brief description on the state-of-the-art in literature concerning the primary topic of Human-Head and Shoulder Segmentation. Additionally, the preliminaries on Skin Detection and Hair Segmentation are reviewed, which gives a brief overview for the algorithm's subtasks of our proposed method on automatic human-head and shoulder segmentation. At the end of Chapter 2 we provide our Basic Methodological Strategy, where the different components of our algorithm are set into relation to each other. In Chapter 3 the first

component Face Skin Silhouette Detection is described and evaluated. In Chapter 4 two Superpixel algorithms are compared and evaluated, since oversampling the input image is a crucial step in processing our proposed algorithm. Chapter 5 describes the last component of our method, followed by an evaluation and discussion regarding the results. Chapter 6 concludes this master thesis and closes with an outlook to future work.

CHAPTER 2

Preliminary

Most literature regarding this topic is treating face detection, face recognition, fake face detection, tracking and feature extraction, so focusing more on the face only rather than the hair and the rest of the body as well. The specifitc problem of automatic human-head and shoulder segmentation is relatively new and was first introduced by Xin et al. [64] in the year of 2011. Since then few papers have been published which will be described in the subsequent Section 2.1. In our approach, we are splitting the problem into different subtasks, such as *Face Skin Detection* and *Hair and Shoulder Detection*, which together compose the human-head and shoulder segmentation. However, if these subtasks are viewed independently and given certain preprocessing steps, they can be applied for different applications as well. Therefore in this Chapter we will describe the current state-of-the-art on Human-Head and Shoulder Segmentation in Section 2.1, followed by the state-of-the-art on Skin Detection in Section 2.2 and Hair Segmentation 2.3. The Chapter closes with the short description of our Basic Methodological Strategy in Section 2.4 of our approach giving an understanding on how these subtasks are combined with each other to achieve a complete strategy on human-head and shoulder segmentation.

2.1 Previous Work on Human-Head and Shoulder Segmentation

The discussion on object segmentation is one of the basic issues in image processing and computer vision. Extensive studies on segmentation algorithms are presented in the literature. Rother et al. [45] presented the GrabCut algorithm to address the problem of interactive extraction of a non-articulated foreground object in a complex environment, whose background cannot be trivially subtracted. The disadvantage is the need of user interaction and only considering color differences. Recently another object detection, recognition and segmentation algorithm has been proposed by Facebook AI

Research (FAIR) [1] based on Deep Learning. Introducing the DeepMask [42] segmentation framework coupled with the SharpMask [43] segment refinement module, they enable the machine vision system to detect and delineate every object in an image. The final stage of recognizing and classifying the object is done by the convolutional network [61].
Segmentation of articulated object, more specific human segmentation is important in dynamic (video) as well as still images. Shu [53] addresses the problem of human detection and tracking in surveillance videos. Thombre et al. [56] uses a simple image segmentation technique for human detection and Kalman filter for human tracking. Human-head and shoulder segmentation on still images has received considerably less attention.

Xin et al. [64] introduced a first approach dealing with this issue in 2011. In their approach an iterative shape mask guided graph cut algorithm with sketch constraints is applied to the oversampled image graph using Watershed Algorithm [59] to get a border that segments the head-shoulder from its background. The limitations stated by the authors are that the algorithm fails when the ground truth is far away from trained shape masks. Furthermore the shape sketch constraint suggests to deal only with a particular hair style.
In Bu et al. [7] the same authors tried a different approach using structural patches tiling to guide the human head and shoulder segmentation. They apply a local structure classifier trained by random forest to the input image in a sliding window manner and then construct an Markov Random Field (MRF) to build a probabilistic mask from the responses collected from the previous stage. They compare their results outperforming GrabCut [45] and achieving similar results as their previously published paper [64].

Jacques et al. [21] propose a head-shoulder contour estimation model for human figures in still images, captured in a frontal pose. The contour estimation is guided by a learned head-shoulder shape model, initialized automatically by a face detector. A graph is generated around the detected face with an omega-like shape, and the estimated head-shoulder contour is a path in the graph with maximal cost. The authors improved their human contour estimation in Jacques and Musse [20] through clusters of learned shape models. However, it is important to emphasize that Jacques et al. [21] and Jacques and Musse [20] try to segment the most omega-like head-shoulder contour, focusing on a well known shape/feature of the human body [20], which means that this contour could include the persons hair, but not necessarily has to.

Sangüesa et al. [47]'s publication is based on the previously mentioned GrabCut [45] algorithm. Computing an initial estimation of the foreground they avoid an manual interaction as input for GrabCut and perform the algorithm for a certain number of iterations. They focus on passport images, which require an almost pixel-perfect segmentation in order to be a valid photo. Their evaluation shows lack of non-uniform backgrounds and was not tested on such challenging scenarios [47].

The last and most recent publication was by Deng and Wu [11] in 2016 and they present a learning-based method for robust head-and-shoulder segmentation results in applications

[1] https://research.fb.com/learning-to-segment/

where the person in the query image is a known prior and the background is non-uniform. In this case a prior knowledge and a portrait image of the person is required to create the head-shoulder object (HSO) and train the method, before predicting on new images of the same subject.

Most of these proposed approaches are concentrating only on color information. In our master thesis we split the problem into different subtasks and consider specially for the hair segmentation texture information as well. Furthermore, the challenge of not having any prior knowledge on the person in the image and the background complexity increases the difficulty we are focusing on in this thesis.

2.2 Previous Work on Skin Detection

Saxen and Al-Hamadi [48] categorizes skin segmentation algorithms into threshold-based, model-based and region-based methods.

Thresholding-based methods are the simplest and most frequently used human skin detection methods, where a fixed decision boundary is defined [55]. For each color space component single or multiple ranges of threshold values are defined. The pixel values of the input image that fall within those predefined ranges are labeled as skin pixels, all the others are defined as non-skin. Liensberger et al. [34] are applying for their online video annotation a combination of YCbCr, normalized RGB and RGB for skin detection. One of the drawbacks of working in the *RGB* color space is that luminance and chrominance cannot be separated. The *RGB* components are highly correlated, so changing the luminance of a given skin patch affect each component.

Transforming from *RGB* into any of the orthogonal color spaces is a linear transformation [13]. All these color spaces separate the illumination component (Y) from the two orthogonal chrominance components (*UV*, *IQ*, *CbCr*). Therefore, unlike the *RGB* color space the location of the skin color in the chrominance components is not affected by changing the intensity of the illumination [13]. The simplicity of the transformation and the invariant properties made these color spaces widely used in skin detection applications [50, 15].

Perceptual color spaces are described by *HSI*, *HSV*/*HSB*, and *HSL*. They separate three components: hue (H), saturation (S) and brightness, also called intensity, value or lightness (I,V, or L). These color spaces are deformations of the *RGB* color cube and are computed by a non-linear transformation. The boundary of the skin color class is specified in terms of hue and saturation. The brightness component I, V and L is often dropped to reduce illumination dependency of skin color. Shaik et al. [50] as well as Platzer et al. [44] used these color spaces in their skin detection approaches.

Commonly used **model-based methods** in literature are Gaussian classifiers or Gaussian Mixture Models (GMMs), which try to approximate the skin-color distribution [25].

Greenspan et al. [16] show a mixture of Gaussians as a robust representation that can accommodate large color variations, as well as highlights and shadows. They trained GMM with two components, where one component captures the distribution of the skin color while the other captures the distribution of the highlighted regions of the skin.

Lee and Yoo [30] compare the performance of a single Gaussian model (SGM) with a GMM of six components. Under controlled illumination condition, skin colors of different individuals in a orthogonal color space cluster in a small region. Hence, in these conditions the skin color distribution can be modeled through an elliptical Gaussian joint probability distribution function (pdf). Once other image conditions have to be considered a SGM is not sufficient and GMMs with multiple components have to be considered. The key idea behind using multiple components is that different parts of the face are illuminated in a different manner and they can be modeled by different components [25].

Lü and Huang [35] propose a skin detection method based on the cascaded adaptive boosting (AdaBoost) classifier, which consists of minimum-risk based Bayesian classifier and models in different color spaces such as HSV (hue, saturation, value), YCgCb (brightness, green, blue) and YCgCr (brightness, green, red). Ma et al. [36] proposed the Semantically Constraint Skin Detection (SCSD) method based on Random Forests. The semantic constraint is based on the dependence between skin pixels and human body parts, to limit the influence of background skin-like pixels. Khan et al. [29] compare their random forest based skin detection approach with other classification learners like Bayesian network, Multilayer Perceptron, SVM, AdaBoost, Naive Bayes and RBF network.

The third possibility is incorporating spatial information, using a **region-based methodology**. A common region-based method used for skin segmentation is Region Growing [48]. The problem with Region Growing is the need of seed points. Abdullah-Al-Wadud et al. [2] use a color distance map and based on this map they generate some skin as well as non-skin seed pixels. Then they grow them to capture the appropriate regions. With this approach they do not generate much noisy segments and do not need any prior training session. Saxen and Al-Hamadi [48] propose a region growing approach computing the seed points by a Bayes approach.

Khan et al. [28] propose a skin segmentation approach using graph cuts. They model the skin segmentation as a min-cut problem on a graph defined by the image color characteristics and a universal seed to overcome the potential lack of successful seed detections. The advantage of their approach is that it is only based on skin sampled training data making it robust to unseen backgrounds.

In this thesis we present a model-based supervised classification learner based on independently decision trees and weighted kNN. We include high-level information of the query image into the training set and show how this improves the skin detection. The data used as training and testing sets were transformed from RGB color space into the

orthogonal color space YCbCr from which the two chrominance channels Cb and Cr represent the feature space.

2.3 Previous Work on Hair Segmentation

Hair plays a significant role in the overall appearance of an individual and many computer vision tasks can benefit from segmented hair. For instance, it provides an important clue for gender classification, since hair styles (including facial hair) of male and female are generally different. Often hair can also facilitate the automatic age estimation of a person, since hair volume, density and color gradually changes (or disappears in case of baldness) with the increase of age, especially for old men and women [62].
For humans hair is a major cue for face recognition as well [65], changes in hairstyle or facial hair can mislead the observer in recognizing faces, suggesting that it could be of advantage to use hair information in recognition to provide a useful cue for identification or at least narrowing possible matches. However, hair appearance and attributes can easily be changed, and therefore should be treated with caution when it comes to identification.
Wang et al. [63] describes another possible application being, people wanting to see whether or not some hair style fits them or not. With the rapid development of internet, online makeup has become more popular and a good hair style identification or search tool is necessary, which makes hair segmentation essential.
Shen et al. [51] presents an approach for the application of automatic facial caricature synthesis where an accurate detection and presentation of hair region is one of the key components as well. Another interesting application is *AutoHair* introduced by Chai et al. [8] reconstructing a 3D hair model from a 2D image. Even in the beauty industry with augmented reality Levinshtein et al. [32] addressed the topic of live hair color augmentation.

Detecting and segmenting hair within images represents a significant challenge due to the diversity of hair patterns and background variability [62] and in the literature very different approaches were presented.

Wang et al. [62] propose a two-tier Bayesian based method for hair segmentation. In the first tier, Wang et al. [62] uses a Bayesian Model by integrating hair occurrence prior probabilities for computing the initial hair seeds selection, which later on in the second tier are used to built the hair-specific Gaussian model. The algorithm is finalized with Mean Shift results to remove holes and spread hair regions. Relying on Mean Shift to fill in the holes, spreading hair regions, for the final segmentation could lead to adding superpixels which contain small hair information and large background areas if the superpixel granularity is not high enough or the boundary coherence not correct. Furthermore depending only on color information narrows the set of images to handling simpler backgrounds.

Aarabi [1] proposes in their paper an automatic hair segmentation method by extracting in a multi-step process various information components from an image, including background color, face position, hair color, skin color and skin mask with region growing. Regions far

away from the face are considered to define a *background color likelihood histogram*. For the hair detection, they obtain an initial guess on the hair information by taking narrow strips above the face and narrow strips on the sides of the face defining the *hair color likelihood histogram*. With these two likelihoods the rest of the pixels are classified. To improve the hair detection cleanup post-procedures are used removing eyebrows, eyes, island region patches, and strands or segments that point upwards.
Since it is assuming hair above the face and on the side the limitations are that the data set which can be used on this approach is constraint to have long hair and a celebrity alike hairstyle at best as their dataset results show. Their description of how they define their background regions is vague *far away from the face*, which concludes into the assumption that a uniform background is expected.

Wang et al. [63] proposes a learning approach, called Compositional Exemplar-based Model (CEM) for hair segmentation. CEM generates an probabilistic mask in the manner of Divide-and-Conquer, which can be divided into a decomposition and composition stage. In the first stage a strong ranker based on a group of weak semantic similarity features is learned. In the second composition stage, a neighbor label consistency constraint reduces the ambiguity between data representation and semantic meaning and then recomposes the hair style using alpha-expansion algorithm. The final segmentation result is obtained by Dual-Level Conditional Random Fields. The approach shows difficulties and hair being confused with the background when shadows occur or color contrast to the background is low. Moreover an exact result of the test image can only be guaranteed if its hair characteristics exist in the training library.

Rousset and Coulon [46] was as far as our knowledge goes the first to introduce frequential information into the process of hair segmentation. Their algorithm is divided into two steps. Firstly performing a raw segmentation based on frequential and color analysis to place markers in hair regions. Secondly a matting process is used to achieve a final hair mask. The crucial limitation about this approach is that it is bound to a particular set of images, where the background is not that high frequential so that the threshold for their frequential map holds for the hair otherwise no hair regions are found in the frequential map. Moreover too small markers could lack on enough color information and lead to bad estimation of the alpha matte.

Ahn and Kim [4] propose a face and hair region labeling semi-supervised spectral clustering-based multiple segmentation approach introducing texture information to improve the object class distinction. For the training dataset they generated superpixels with watershed algorithm [59] on the frontal-view face images to extract color (in *Lab* color space) and texture with Leung-Malik filter bank [31] features.

Liang et al. [33] provides a hair segmentation solution combining the outputs of a color camera and a depth camera. With the additional depth map a face mesh is computed with which the head region skin and hair can be segmented. However, here an additional depth camera is needed which in practice is not commonly used and expensive.

In this thesis we present in Chapter 5 a method to detect and segment hair and shoulder

automatically using the results of the Face Skin Silhouette Detection (see Chapter 3) and the oversampled image (see Chapter 4) to built a hair, shoulder and background model based on color, texture and location for the individual image to classify the rest of the unknown superpixels in between. Similar to previously mentioned approaches in literature such as [46], [62], [1], [24], [65] we concentrate on frontal-view face images, where the subject is not bald nor semi-bald.

2.4 Basic Methodological Strategy

Our methodological approach to accomplish the expected result of a background reduction in a digital frontal-view face image comprises the following subtasks visualized in Fig. 2.1:

Following the enumeration of the subtasks in Figure 2.1, at first, the eyes in the input image Figure 2.1(1) are detected to place the control points for an Active Contour Model (ACM) [27] in Figure 2.1(2). The shape mask identified with the blue contour line in Figure 2.1(2) is the initial mask for ACM. The purpose of this rough separation into foreground and background is to segregate most of the background out of the image resulting in an incomplete background mask (3a) and a foreground mask (3b) with spurious segments including the subject. For images with a salient object and a simple, noiseless and uniform background such a segmentation through ACM could be sufficiently enough to produce an acceptable segmentation result, but for our problem statement of dealing with complex images regarding foreground and background this procedure leads to a foreground mask with erroneous regions.

From these rough background and erroneous foreground masks automatically labeled data is extracted to improve the performance of our face skin classifier (in Figure 2.1(4)). The skin detection (described in Chapter 3) determines pixels, which are certainly part of the foreground.
To evaluate this subtask of our face skin detection algorithm, we compare quantitative results with other skin detection strategies measuring the performances in terms of True Positive Rate and False Positive Rate and comparing qualitative results with existing skin detection explicit thresholding methods.

With the results provided by ACM and our Skin Detection algorithm the image is subdivided into a trimap: pixels which are certain to be background (6a), pixels which are detected as skin (and all areas surrounded by skin like eyes, eyebrows, mouth) classified as foreground (6c), and pixels in between labeled undefined (6b). To classify this last undefined area of pixels correctly the image is oversampled into superpixels in subtask (5), to extract per superpixel color and texture information and their relative location in the image. With these per superpixel informations, hair, shoulder and background models can be characterized, to determine in the final stage whether the remaining superpixels correspond to the foreground or background. The final boundary mask describes the result shown in Fig. 2.1-(7).

To evaluate the superpixel segmentation we compare two state-of-the-art algorithms,

SLIC (Simple Linear Iterative Clustering) by Achanta et al. [3] and GS04 (Efficient Graph-Based Image Segmentation) by Felzenszwalb and Huttenlocher [14], discussing the granularity with respect to our problem statement (see Chapter 4).

The final stage of the methodology, hair and shoulder segmentation, is described in Chapter 5 and to assess the resulting human-head and shoulder segmentation quantitative as well as qualitative evaluations were conducted, measuring the consistency of segmentation results with manually labeled ground truth in terms of the overlap ratio. This evaluation criterion quantifies the error of labeled foreground containing background information and vice versa.

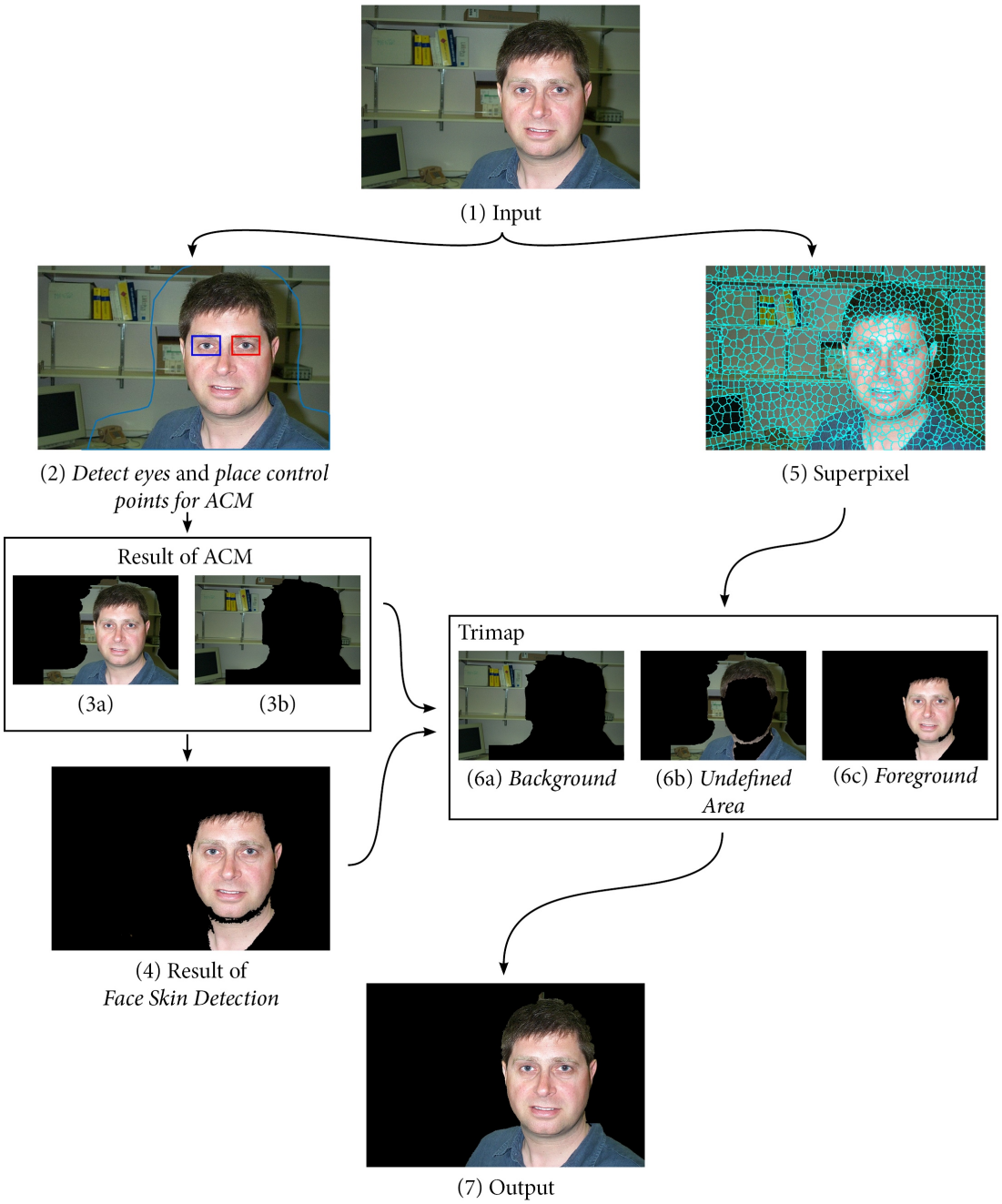


Figure 2.1: Overview of our approach: (1) Input image. (2) Detect eyes and placing control points for a initial mask (see blue line). (3*) ACM Results (3a) *Foreground* & (3b) *Background*. (4) Face Skin Detection. (5) Superpixel Segmentation. (6) Trimap: Decision Procedure for ambiguous pixels in undefined area considering their color, texture and location information. (7) Output Image.

CHAPTER 3

Face Skin Silhouette Detection

Skin detection in general is the process of finding skin colored pixels and regions in an image or a video [13]. It is the process of separating skin and non-skin pixels. Skin detection is an important feature for several computer vision applications, such as face detection and tracking [9] or hand detection and tracking [12], which is widely used for navigation and object manipulation in Virtual Reality (VR) or Augmented Reality (AR) [23]. Other examples are the retrieval of humans in databases and the Internet, automatic annotation, archival and retrieval [34], and content filtering for parental control software or criminal investigations [44]. In this master thesis we concentrate on images with frontal-view face images and look at skin detection as a preprocessing step of head and shoulder segmentation. Therefore our main interest lies on the classification of skin pixels around the silhouette of the face, neck or possible shoulders of humans.

3.1 Technical Specification

For computer vision systems, skin detection is prone to many challenges and still an open problem, while for the human visual system skin detection is easy. Spillmann and Werner [54] describes the human perception with an example of seeing a blue ball, were we all can agree in that the ball is perceived blue as whole, and not as a ball having blue patches and some other color patches produced by differences in illumination. Furthermore, the human visual system can dynamically adapt to varying illumination conditions, so it can preserve the actual color of the object [25]. In literature this is called color constancy or chromatic adaptation.

Most of the literature on human skin detection has focused on using color information, which can be a challenging task as the skin color in images is sensitive to various factors such as illumination, camera characteristics, ethnicity and skin-alike background [25].

We specify the technical specifications for the input image as follows, which will be referenced in the subsequent sections of this chapter:

1. In this thesis we are focusing on frontal-view face images, which means that the *pose* of the person's head and shoulders captured has to face the camera, so the rotation of the head shall be less than $+/-5$ degrees from frontal in every direction roll, pitch and yaw.
2. Based on human anatomy we know that in general a human being has two eyes, one nose, one mouth and two ears. Hence, in a static image where the person is looking towards the camera in a frontal-view pose, these *facial features* will be visible, except for the ears which could be occluded by hair. As an additional requirement, the persons eyes must be open; Closed or covered eyes are not accepted.
3. Since face skin detection is considered here as a preprocessing step, the focus lies on finding a correct silhouette of the subjects face and neck. If the pixels around the silhouette are correctly classified, the remaining pixels inside the silhouette can be labeled as foreground and everything outside as background.

3.2 Face Skin Detection based on Classification Learners

We propose a novel skin detection algorithm based on classification learners. In pattern recognition, classification is considered an instance of supervised learning, e.g. learning where a training set of correctly identified observations is available. In literature there are a number of algorithms including [37]: Linear Classifiers, Support Vector Machines (SVM), Kernel estimation like k-Nearest Neighbor (kNN), Boosting (meta-algorithms), decision trees and neural networks (NN). These algorithms were studied during the master thesis and for further information on why we decided on weighted kNN and decision trees for this particular problem we refer the reader to our published technical report [39].

The novelty of the proposed approach lies in our improvements on the training set of the kNN and decision trees classifier (see Section 3.2.3).

3.2.1 Recall: Decision Trees

Decision trees are characteristic in having fast prediction speed[1], small memory usage[2] and being easy to interpret. A disadvantage can be that they have low predictive accuracy and tend to overfit, if the depth of the splits is not pruned to a maximum number of splits [22]. We decided on a decision tree with a maximum number of 100 splits, which could lead to overfitting on the training set. Since we are improving our classification learner as described in Section 3.2.3 with information of the input image itself a detailed decision tree is more suiting to classify the remainders of the input image correctly. In the following evaluation we refer to this methodology by *tree*.

[1] Speed: Fast 0.01 sec.; Medium 1 sec.; Slow 100 sec.
[2] Memory: Small 1MB; Medium 4MB; Large 100MB

3.2.2 Recall: Weighted k-Nearest Neighbor (kNN)

Nearest Neighbor classifiers are characteristic in having slow to medium prediction speed[1], medium memory usage[2] and being harder to interpret compared to decision trees. They typically have good predictive accuracy in low dimensions. As dimensionality increases, the distance to the nearest data point approaches the distance to the farthest data point, which might lower the prediction accuracy. In the k-Nearest Neighbor (kNN) algorithm categorizing a query point is based on its closest k neighbors in the training examples. In the weighted kNN the distances to the neighboring points are weighted. Choosing a high number of neighbors can be time consuming to fit. For the evaluation in Section 3.3 a number of 10 neighbors was defined and a distance weight of squared inverse. It is referenced by *kNN*.

3.2.3 Face Skin Classification Learner (FSCL)

To improve the performance of classification learners the training data is extended with automatically extracted sample information of the query image. A series of preprocessing steps were performed on the input image to extract pixel information to be included into the training set.

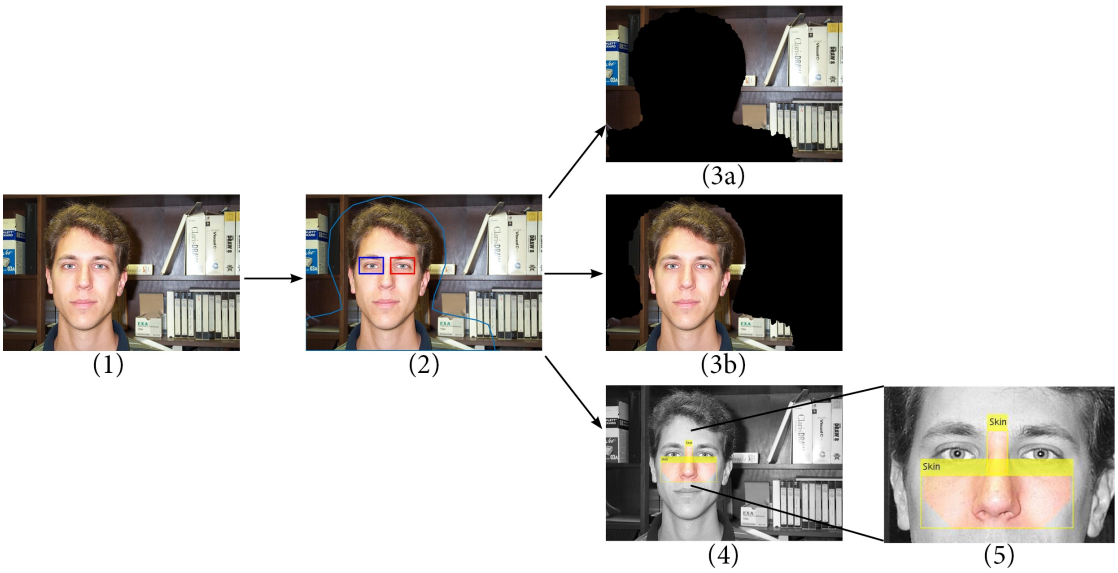


Figure 3.1: Overview of the preprocessing steps: (1) Input image. (2) Detect eyes and placing control points for the initial ACM mask (see blue line). (3*) ACM Results (3a) *Foreground* & (3b) *Background*. (4) Extracting skin pixels (in color) (5) Zoomed into the selection of the extracted skin information.

Following the enumeration of the subtasks in Figure 3.1, at first, the face and eyes in the input image Figure 3.1(1) are detected by Viola-Jones [60]. Viola–Jones requires full view frontal upright faces. Thus in order to be detected, the entire face must point

towards the camera and should not be tilted to either side. These requirements are met, since we are focusing in this thesis on frontal-view face images and as described in our technical specification 1 the person's face captured by the camera is in a frontal-view pose, leading to technical specification 2 as a trivial result showing all facial features. With the face and eyes location and dimension the control points for an Active Contour Model (ACM) [27] in Figure 3.1(2) are placed. The shape mask identified with the blue contour line in Figure 2.1(2) is the initial mask for ACM. The purpose of this rough separation into foreground and background is to segregate most of the background out of the image resulting in an incomplete background mask (3a) and a foreground mask (3b) with spurious segments including the subject. The last preprocessing step is the extraction of human skin information of the query image, shown in Figure 3.1(4). With the same Viola-Jones algorithm [60], but different Haar-like features the nose of the person is found in the image underneath the eyes location. With this information skin pixels are extracted from the region between the eyes location and the nose bounding box, represented as the two skin boxes in Figure 3.1(4). In the following evaluation we refer to our two improvements of the classification learners by *tree-FSCL* and *kNN-FSCL*.

3.2.4 Color Space YCbCr for Skin Detection

As color space for training both classification learners the orthogonal color space YCbCr was chosen, since orthogonal color spaces like YCbCr separate the illumination component (Y) from the two orthogonal chrominance components (CbCr). Unlike the RGB color space the location of the skin color in the chrominance components is not affected by changing the intensity of the illumination [13]. According to Elgammal et al. [13] the skin color of different ethnicity groups almost co-locates in the chrominance channels.

Observing the histograms of the publicly available *UCI* database (see detailed description in Section 3.3.1) containing skin and non-skin pixels once in RGB color space (Figure 3.2) and in YCbCr color space only considering the chrominance components Cb and Cr (Figure 3.3) can be observed that in the RGB the non-skin pixels overlap completely with the skin pixels making the correct classification harder. For the YCbCr color space only a smaller overlap can be observed for this particular database, which makes it more suiting for a correct classification.

Figure 3.4 shows the incorrect classified pixels after training a decision tree with *UCI* dataset once in RGB color space and once in YCbCr color space only considering the chrominance components. In orange are the false positive and in blue the false negatives. The decision tree trained in CbCr color space classifies 0.14% less incorrectly than the decision tree trained in RGB color space regarding the *UCI* database.

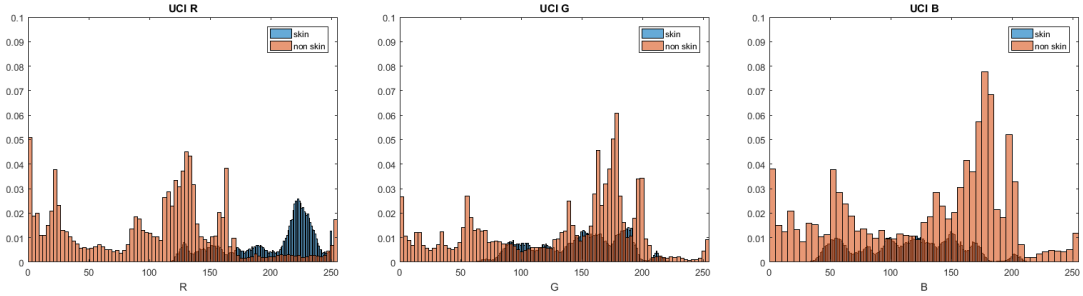


Figure 3.2: 1-D histograms of skin vs. non-skin pixels of the *UCI* database in RGB color space.

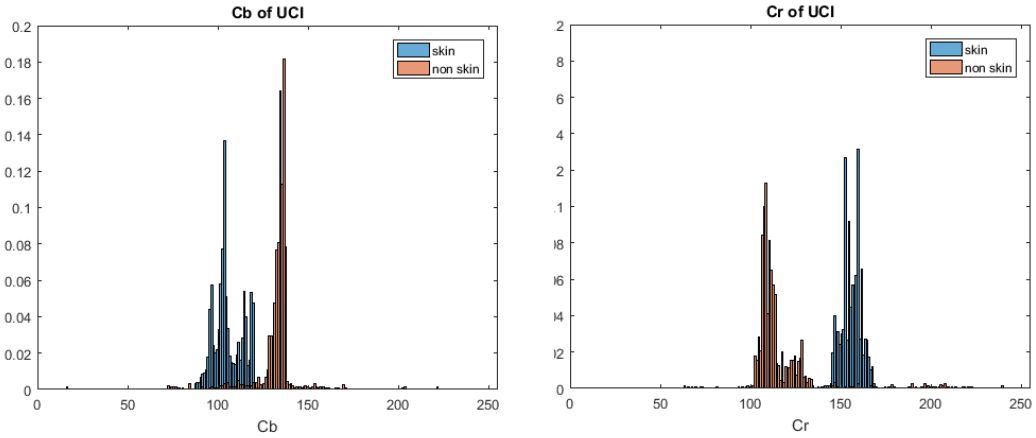


Figure 3.3: 1-D histograms of skin vs. non-skin pixels of the *UCI* database considering the chrominance components of the YCbCr color space.

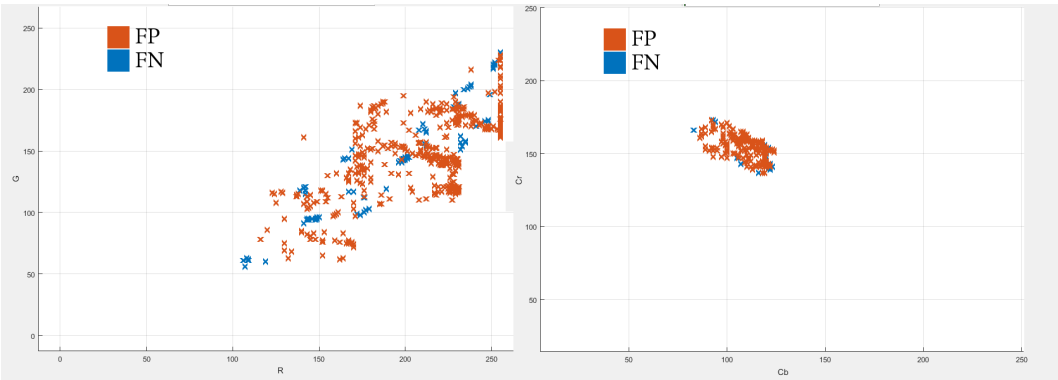


Figure 3.4: Incorrectly classified pixels from decision tree. Left: Result of trained decision tree in RGB color space. Right: Result of trained decision tree in YCbCr color space. In orange are the false positive and in blue the false negatives.

3.3 Results and Evaluation

The proposed approaches based on the classification learners decision tree and weighted kNN and their improvements were implemented in MATLAB[3]. In the following qualitative and quantitative evaluation we compare our proposed approaches *tree* and *kNN* and their improvements *tree-FSCL* and *kNN-FSCL* with skin detection based on *explicit thresholding in the YCbCr Color Space* [13] (thresholdYCbCr), *HSV Color Space* [15] (thresholdHSV) and *RGB Color Space* [15] (thresholdRGB).

In a further analysis as evaluation criterion, only the silhouette of the ground truth is considered (see Figure 3.5). As described in the technical specification 3 the region of interest of our facial skin detection lies on the silhouette of the persons face and neck. If the pixels around the silhouette are correctly classified then the rest inside the silhouette can be labeled as face and everything outside the silhouette as background.

(1) (2) (3)

Figure 3.5: Focus in the evaluation of skin detection on the silhouette of the person. (1) Original image. (2) Ground truth. (3) New silhouette ground truth.

Representative sample images from different databases were selected to demonstrate the performance and limitations of the proposed approach. Regarding quantitative evaluations the segmentation results of the approaches were compared against the ground truth and the silhouette ground truth (see Figure 3.5). In the context of skin classification,

- *true positives* are skin pixels that the classifier correctly labels as skin.
- *true negatives* are non-skin pixels that the classifier correctly labels as non-skin.
- *false positives* are non-skin pixels that the classifier erroneously labels as skin.
- *false negatives* are skin pixels that the classifier erroneously labels as non-skin.

The goal of a good classifier is to have low false positive and false negative rates. As in any classification problem, there is a trade-off between false positives and false negatives [13]. Having a soft class boundary the false negative rate is low and the false positive rate is high, which results in a high recall value. Having a tighter class boundary the false negatives are high and the false positives low. This normally results in a higher precision

[3]MATLAB: `https://de.mathworks.com`

value.
For a more detailed evaluation and survey of results we allow to refer the reader to our published technical report [39].

3.3.1 Databases

Experiments are conducted using the following public datasets which all except for the last one provide a ground truth. The databases were transformed from RGB color space into the orthogonal color space YCbCr, from which the two chrominance channels Cb and Cr represent the two-dimensional feature space.

It is important to mention that the primary focus is on images, where the face can be easily found with state-of-the-art face detection algorithms, so the subject in the image is not occluded and face and shoulders are facing the camera.

- *UCI [6]:* is collected by randomly sampling B,G,R values from face images of various age groups (young, middle, old), ethnicity groups (white, black, and Asian), and genders obtained from FERET database and PAL database. The dataset provides ground truth and contains 245.057 pixel entries (50.859 skin and 194.198 non-skin).

- *Pratheepan [55]:* is collected randomly from Google and images are captured with a range of different cameras, using different color enhancement, under different illuminations, variation of age (young, middle), ethnicity groups (white, Asian), and genders. The database provides ground truth and contains 32 face images.

- *CALTECH* [4]*:* this frontal face dataset is collected at California Institute of Technology, capturing 27 people under different light conditions, facial expression, ethnicity groups (mostly white and Asian), gender and complex backgrounds. It provides images under different conditions with a complex background, where the orientation of the head and shoulders is facing the camera according to the defined criteria we are focusing on in this thesis. The database does not provide any ground truth. Therefore, for a small set of images ground truth was generated manually and those samples were used in qualitative evaluations.

3.3.2 Evaluation of FSCL

In this subsection we are discussing quantitative and qualitative results concerning the proposed FSCL approach and compare it with state-of-the-art algorithms. For the evaluation we are using *UCI* database as training set for the classification learners *tree* and *kNN*. As described in Section 3.2.3 *tree-FSCL* and *kNN-FSCL* include in the training phase information of the input image and the *UCI* database. Both quantitative results

[4] Collected by Markus Weber at California Institute of Technology `http://www.vision.caltech.edu/html-files/archive.html`

in Tables 3.1 and 3.2 are realized with *Pratheepan* as testing set. In the first Table 3.1 the complete provided ground truth has been considered. In the second Table 3.2 the results are regarding only the correct classification around the silhouette of the subjects skin region. Some qualitative results are provided in Figure 3.6.

The results concerning the complete ground truth of skin are shown in Table 3.1. The best performance regarding accuracy, precision and F1 measure is our *tree-FSCL*. All classification learners outperfom the explicit thresholding methods regarding the *Pratheepan* database as testing set. The explicit thresholding methods *thresholdYCbCr* and *thresholdRGB* are prone to generally classify more pixels as skin, leading to a high value of true positives but also false positives. This can also be observed in the precision value, which considers the false positive rate in its calculation.

Approach	Accuracy	Precision	Recall / TPR	FPR	F1
tree-FSCL	**0.934**	**0.852**	0.848	**0.052**	**0.841**
kNN-FSCL	0.926	0.818	**0.869**	0.067	0.831
tree	0.908	0.796	0.842	0.080	0.797
kNN	0.910	0.794	0.860	0.083	0.807
thresholdYCbCr	0.690	0.348	0.774	0.356	0.450
thresholdHSV	0.738	0.319	0.419	0.215	0.323
thresholdRGB	0.695	0.330	0.657	0.320	0.409

Table 3.1: Evaluation on the testing database *Pratheepan* concentrating on the complete ground truth.

The results of Table 3.2 are evaluating the classification only around the silhouettes ground truth. Our proposed classification learners do not outperform the explicit thresholding methods in Recall and F1-score even though the same testing database of *Pratheepan* has been used for both evaluations (Tables 3.1 and 3.2). Recall is higher for *thresholdYCbCr* and *thresholdRGB*, because both find more skin pixels but as a drawback also categorize a large number of background pixels as skin (see FPR). This could have great negative impact on the further process of segmenting out the background from the person (as can be observed more clearly in the qualitative results in Figure 3.6). Regarding accuracy and precision the supervised classification learner based on decision tree *tree-FSCL* outperforms the other algorithms.

To give a further comparison, in the latest survey of skin-color modeling and detection methods by Kakumanu et al. [25], the authors compare skin detection strategies and their performance in terms of the *true positive rate (TPR)* and *false positive rate (FPR)*. Obviously it is difficult to compare these different published methodologies, since there is no uniform benchmark dataset on skin detection like there is on general image segmentation and boundary detection (Berkeley Segmentation Dataset and Benchmark [38]). Therefore we have to keep in mind that the results listed in this report are all concerning their own dataset with a respective ground truth.

Approach	Accuracy	Precision	Recall / TPR	FPR	F1
tree-FSCL	**0.797**	**0.799**	0.778	0.205	0.772
kNN-FSCL	0.788	0.771	0.801	0.245	0.770
tree	0.764	0.757	0.778	0.264	0.743
kNN	0.765	0.753	0.793	0.274	0.751
thresholdYCbCr	0.698	0.64	**0.914**	0.515	0.745
thresholdHSV	0.720	0.754	0.600	**0.181**	0.644
thresholdRGB	0.775	0.732	0.882	0.333	**0.789**

Table 3.2: Evaluation on the testing database *Pratheepan* concentrating on the silhouette as ground truth.

The best performing algorithms regarding the quantitative results listed in the report, show a confidence value of around 88.5%-99.4% TPR and 10%-15.5% FPR. In our report regarding the *Pratheepan* dataset on the complete ground truth (see Table 3.1), we can observe that for *tree-FSCL* a 84.8% TPR is reached, which is for a small margin below the state-of-the-art results reported in the survey, and 5.2% FPR is achieved, which shows better performance.

For the qualitative examples illustrated in this thesis we selected images with a variety of different skin tones, background and illumination to give a good representation on the tested samples. Observing the first and second examples, the face is illuminated from the side causing a shadow in the background and different skin tone patches in the face of the subject. For the simple explicit thresholding algorithms these areas are difficult to distinguish and classify correctly. The results of the classification learners are in these two samples better. Furthermore, observing the difference between *tree-FSCL* and *tree* a noticeable improvement can be detected regarding the reduction of false positives. Looking at the third example, the results of the classification learners are very similar, still outperforming the explicit thresholding methods.

It can be concluded that our novel supervised skin classifier improves results significantly when we are dealing with complex backgrounds, different ethnicities and different illumination conditions. The simple explicit thresholding methods and the classification learners *tree* and *kNN* have problems distinguishing between skin-alike pixels in the background and actual skin pixels of the person since no contextual information is available. All three examples demonstrate the typical behavior of *tresholdYCbCr* and *thresholdRGB* classifying more pixels as skin, leading to a high true positive and false positive rate.

Allowing too much or too little light during exposure makes images darker or brighter, respectively changing the natural tone of skin. A color space such as YCbCr allows to compensate this problem by splitting color into the luminance and chrominance components. In Figure 3.7 an example of over- and underexposed image can be seen, where the *thresholdYCbCr* results are spurious not finding most of the skin pixels. Using

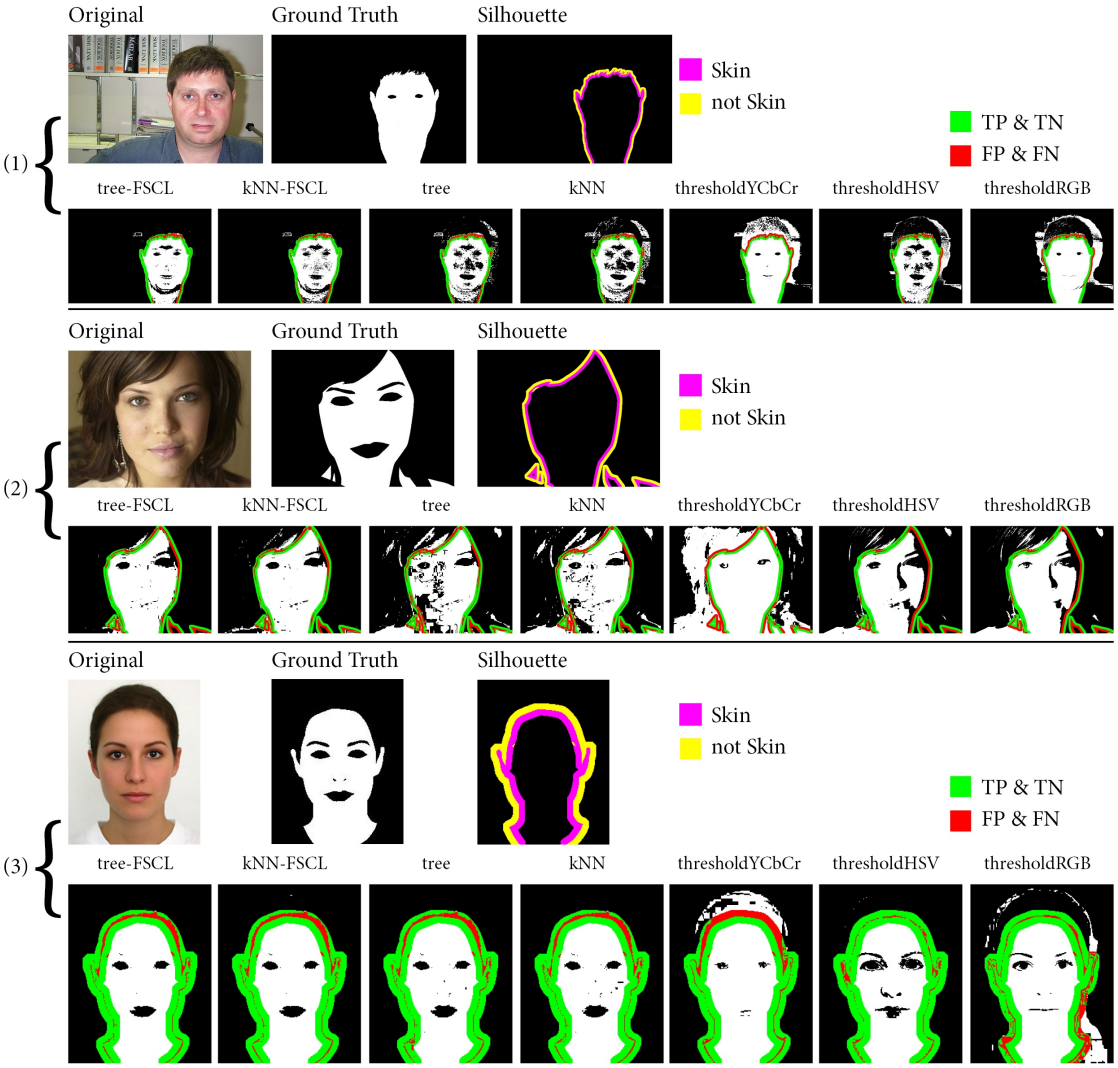


Figure 3.6: Qualitative Examples: image (1) is from *CALTECH* database and images (2),(3) from *Pratheepan*. White pixels are skin, black non-skin and around the silhouette green represent all true positives (TP) and true negatives (TN) and red all false positives (FP) and false negatives (FN).

the idea of classification learners in particular looking at *tree* the results are even worse, but after adding high-level information (skin pixels and background pixels of the input image) in *tree-FSCL* the results improve but having still erroneous regions.

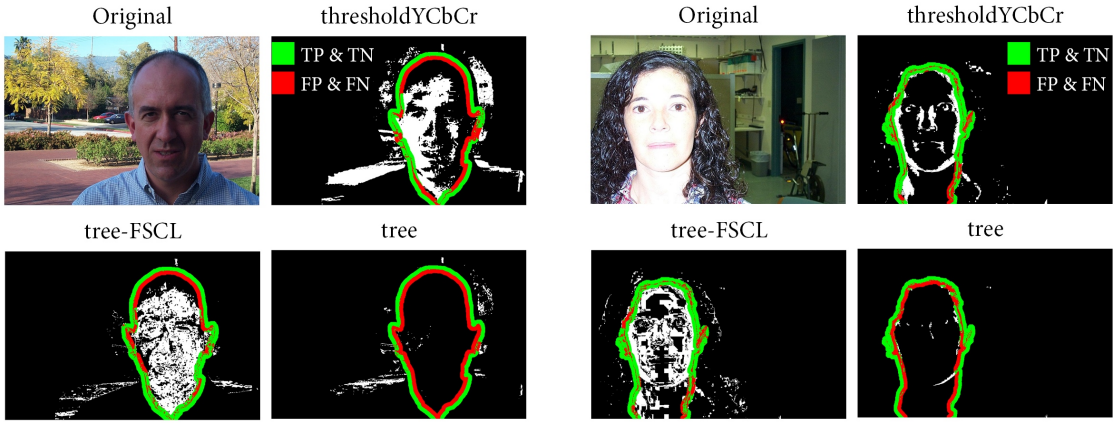


Figure 3.7: Examples of a under- and overexposed image where the results of *thresholdYCbCr* and *tree* fail completely. *Tree-FSCL* improves the skin detection but not sufficiently enough.

3.4 Discussion

In this master thesis we concentrate on frontal-view face images and look at skin detection as a preprocessing step of an automatic human-head and shoulder segmentation. Our main interest lies on the classification of facial skin pixels around the silhouette of the face and neck of humans, since everything in the face skin silhouette can be defined as foreground (skin) as well.
We present a novel model-based approach using classification learners with supervised learning, called Face Skin Classification Learner (FSCL). The proposed solution is based on independent pixel classifiers, namely weighted kNN and decision trees. Both classifiers are trained from automatically labeled data and extended by using Viola-Jones eyes and nose detectors and Active Contour Model (ACM) to extract sample pixels of both skin and non-skin classes.
Evaluations on multiple datasets with frontal-view face images were discussed, and results were compared with explicit thresholding methods. Furthermore we discussed the results of skin detection strategies summarized in the survey report by Kakumanu et al. [25] measuring the performances in terms of true positive rate (TPR) and false positive rate (FPR). The evaluation shows improvements over several baselines and is above the average of the best performing state-of-the-art algorithms regarding FPR. In our particular case, it is more important to have a low FPR rather than low false negative rate, since missing skin pixels can be compensated in the following subtasks of our methodology through the oversampled image into Superpixels or the closing morphological operation and filling of holes. When it comes to the falsely classified background pixels as foreground, they are not removed in any post-processing step.
Including information of the input image into the training set and applying FSCL on the remainder of the image allows the reduction of false positive detections significantly and

the classification results around the silhouette become more reliable.
Since we are considering color as single information, difficulties are visible when unnatural skin tones occur through shadows, over and under exposure or color bleeding (the colored reflection of indirect light from a nearby object).

CHAPTER 4

SLIC and GS04 Superpixel Comparison

Achanta et al. [3] describes superpixel algorithms as grouping pixels into perceptually meaningful atomic regions which can be used to replace the rigid structure of the pixel grid. According to this definition, the idea of superpixels is to capture the image redundancy, provide convenient primitives from which image features can be extracted and reduce complexity of subsequent image processing tasks.

In our case this is a relevant pre-processing step to enable an extraction of initial hair, shoulder and background information to build a representative model for the particular image. Per superpixel the color, texture and relative positions to each other is stored to further analyze whether a certain superpixel is part of the background or foreground.

Algorithms for generating superpixels can be broadly categorized as either **graph-based** or **gradient-ascent-based** methods. In graph-based methods each pixel is treated as a node and the similarity between two neighbors define the edge weights. Well known algorithms which were used in the past are: Normalized Cuts Algorithm by Shi and Malik [52] and GS04 by Felzenszwalb and Huttenlocher [14]. Gradient-ascent-based algorithms start with a rough clustering of pixels and iteratively refine the clusters until some convergence criteria is met to form superpixels. Some examples are Mean Shift by Comaniciu and Meer [10], Quick Shift by Vedaldi and Soatto [58], Watershed Approach by Vincent and Soille [59] and most recent one SLIC (Simple Linear Iterative Clustering) by Achanta et al. [3].

For our purpose we chose to work with SLIC, since in the current literature it is defined as state-of-the-art [5] and GS04, since it considers the topology creating very irregular sizes and shapes and has similar computational time compared to SLIC [3]. In the following Sections we explain both superpixel algorithms briefly and compare them in relation to our problem statement.

4.1 Technical Specification

The desired properties of superpixel algorithms are as follows [3]:

1. The idea of superpixels is to group pixels into meaningful regions, simplifying the redundancy and homogeneous regions into superpixels to compute image features and to greatly reduce the complexity. Therefore, most importantly the superpixels should *adhere well to image boundaries*.

2. When a superpixel algorithm is used as pre-processing step to reduce complexity, the algorithm should be fast, memory efficient and simple to use.

4.2 Recall: SLIC (Simple Linear Iterative Clustering)

Among all the superpixel algorithms, the simple linear iterative clustering (SLIC) method is widely adopted due to its practicality and performance (see evaluation results in Fig. 4.3). SLIC is an adaptation of the k-means algorithm, generating compact and regular-sized superpixels by clustering pixels located close to each other based on their color similarity and spatial information. For this it uses a five-dimensional space, namely *labxy*, where *lab* represents pixel color values in the CIELAB color space which is considered both device independent and suitable for color distance calculations, and *xy* represents the coordinates for pixel position. The reason for SLIC having a linear complexity compared with the original k-means algorithm is limiting the size of search region to a constant distance measure, instead of comparing each pixel with every cluster center. Results of SLIC can be observed in Fig. 4.1, where the only parameter is the total number of superpixels for a particular image.

(a) image_0001 from *CALTECH* Database (b) image_0143 from *CALTECH* Database

Figure 4.1: Results of SLIC algorithm with different resolution size of superpixels $N = \{3000, 1500, 500\}$.

4.3 Recall: GS04 (Efficient Graph-Based Image Segmentation)

GS04 [14] is based on selecting edges from a graph, where each pixel corresponds to a node in the graph and neighboring pixels are connected by undirected edges, where the weights on each edge measures the dissimilarity between pixels. Two quantities are compared at the boundary of two regions: one based on intensity differences across the boundary and the other based on intensity differences between neighboring pixels within each region. So, the intensity differences across the boundary of two regions are perceptually important if they are large relative to the intensity differences inside at least one of the regions. To control the size of the components a constant parameter k is defined beforehand by the user. A larger k causes a preference for larger components, however, parameter k is not a minimum component size. For smaller components a strong evidence for a boundary, so a sufficiently large difference between neighboring components, is required. This way however, it does not offer an explicit control over the amount of superpixels or their compactness. This can be observed in the total number of resulting superpixels in Figure 4.2.

Compared with SLIC, for GS04 both fine detail and larger structures are perceptually important, leading to producing superpixels with very irregular sizes and shapes. Figure 4.2b shows that the segmentation preserves small regions such as hair strands which are not considered in the regular atomic superpixels generated with SLIC.

(a) Leading to a number superpixels for image_001 from *CALTECH* Database: 2450 (top), 1905 (middle) and 871 (bottom).

(b) Leading to a number superpixels for image_0143 from *CALTECH* Database: 2374 (top), 1672 (middle) and 665 (bottom).

Figure 4.2: Results of GS04 algorithm with different k values 50 (top), 100 (middle) and 500 (bottom).

4.4 Comparing SLIC and GS04

The desired properties of superpixel algorithms are most importantly the adherence of boundaries (see technical specification 1). How well a method adheres the contours

can be measured with the quantitative evaluations of the boundary recall and the under-segmentation error. The first evaluation describes the fraction of the ground truth edges that fall within at least two pixels of a superpixel boundary. The second measures the amount of superpixel "leak" for a given ground truth region. The third important quantitative evaluation is speed. When superpixels are used, one of the desired properties are to reduce computational complexity as a pre-processing step (see technical specification 2). Superpixels should be fast to compute, memory efficient, and simple to use, so that they improve the following steps of a method.
In the publication of Achanta et al. [3] a number of superpixel algorithms where compared regarding these three criteria on the Berkely database [38]. The results presented show that superpixels generated by SLIC and GS04 demonstrated the best boundary recall performance, which means that very few true edges were missed. Regarding the under-segmentation error, SLIC outperforms the other methods and for the computational time required to generate superpixels SLIC is the fastest, followed closely by GS04 outperforming the others. The complexity of SLIC is $O(N)$ linear in the number of pixels N, irrespective of the number of superpixels, and GS04 is $O(N\log N)$ complex. According to Achanta et al. [3] SLIC is also more memory efficient than GS04.

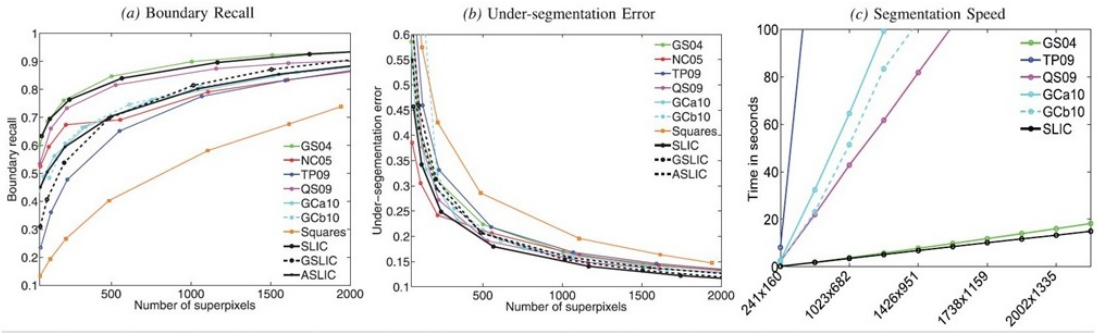
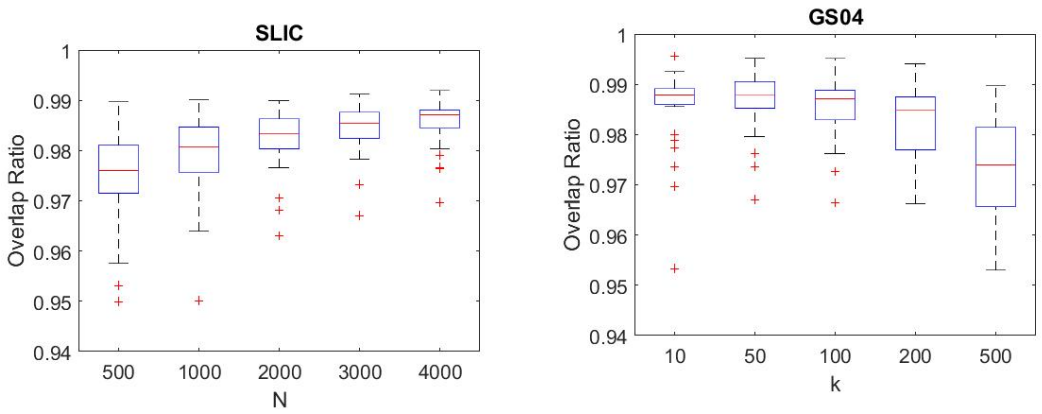


Figure 4.3: Quantitative evaluation measurements from Achanta et al. [3]: The SLIC and GS04 algorithm outperform most of the other State-of-the-Art approaches in (a) boundary recall, (b) under-segmentation error and (c) speed.

When it comes to our problem of a correct human-head and shoulder segmentation the focus lies on the correct boundary detection around the silhouette of the person. So the evaluation criterion is measuring the error when a superpixel contains both foreground and background information. This criterion can be measured with the overlap ratio [64]:

$$overlap = \frac{Ground \cap Segment}{Ground \cup Segment} \tag{4.1}$$

where *Ground* is the image ground truth, and *Segment* is the result of the superpixel algorithm. This evaluation allows to tell how big the error is when supposedly all superpixels are correctly assigned and labeled with background or foreground. Experiments



(a) For different number of superpixels N.

(b) For different values of k. The bigger k the bigger superpixel components can get.

Figure 4.4: Overlap Ratio for both Superpixel Algorithms SLIC and GS04 on 30 images of *CALTECH* Database and *FEI Face* Database.

are carried out on 30 frontal-view images from two datasets CALTECH [1] and FEI Face Database [2] with various hair styles and different skin colors as well as background. It must be kept in mind that a manual segmented image as these contain errors as well, specially because of low contrast around the boundary, single hair strands standing out or artifacts due to Bayer filter interpolation or demosaicing. Experiments were conducted with both superpixel algorithms regarding the overlap ratio in relation to the number of superpixel in Figure 4.4. It can be observed that the more superpixels the better the boundary is preserved, but for SLIC at $N = 2000$ and GS04 at $k = 100$ the margin gets smaller. Even for GS04 at $k = 10$ forcing smaller components and therefore more superpixels (see Figure 4.5 and relation of k and number of superpixels), the overlap ratio is even worsen for a small margin compared to $k = 50$.

Allowing large components in the GS04 superpixel segmentation leads to a smaller number of superpixels as can be observed in Figure 4.5. The larger k is selected the smaller the margin of number of superpixels gets. Important in our case is a trade off between a good oversampling of the image to further extract information per superpixel and adhere the boundaries around the silhouette of the person. In Figure 4.6 an example is demonstrated for a too low granularity of superpixels. For the SLIC algorithm at $N = 500$ as well as GS04 at $k = 500$ the hair and background is incorrectly merged into one superpixel. Having higher granularity and therefore more superpixels the hair and background are separated correctly into different ones.

On the one hand, it can be concluded that a too low granularity does not consider the

[1] Collected by Markus Weber at California Institute of Technology `http://www.vision.caltech.edu/html-files/archive.html`

[2] From the Department of Electrical Engineering in Brazil `http://fei.edu.br/~cet/facedatabase.html`

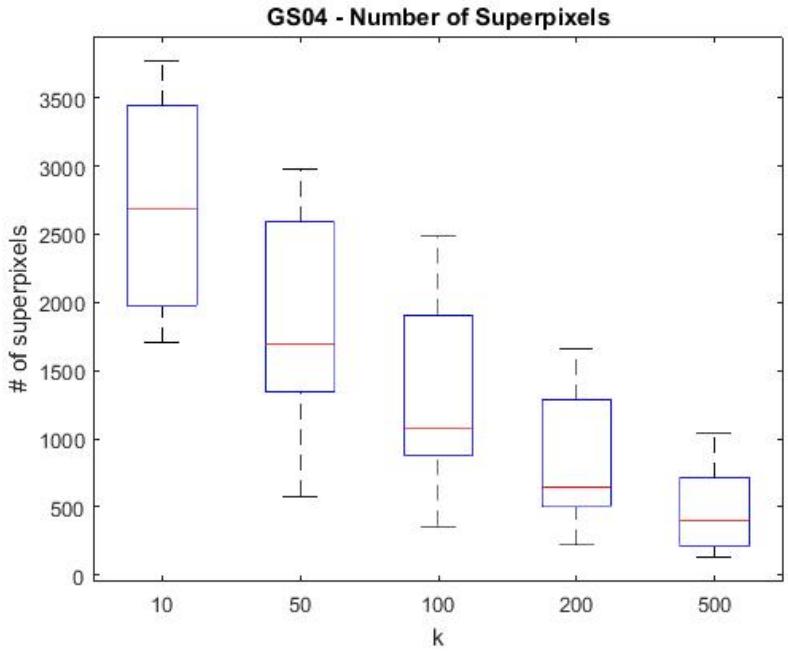


Figure 4.5: GS04 parameter $k = \{10, 50, 100, 200, 500\}$ to control the size of the components. Larger k causes a preference for larger components and therefore smaller Number of Superpixels.

level of boundary adherence we desire for this particular problem and a more detailed superpixel segmentation is needed. On the other hand, having a too high granularity hinders a meaningful per superpixel extraction of information, to enable a correct hair and shoulder segmentation in the following. With SLIC superpixels a better comparison of similarity is guaranteed since the result of SLIC are regular compact superpixels in matters of size (total number of pixels), but with GS04 the topology is considered leading to irregular shaped superpixels taking into account non-rigid objects such as hair with e.g. outstanding hair strands.

4.5 Discussion

We compared and evaluated both superpixel algorithms SLIC and GS04 considering the defined technical specifications and discussed the differences regarding the oversampled output image. SLIC allows to specify the amount of superpixels, is fast in computation and returns compact regular sized superpixels, which is often desired because their bounded size and few neighbors form a more interpretable graph, allowing to extract more locally relevant features [3]. However, compactness comes at the expense of boundary adherence, where GS04 shows better results regarding overlap ratio and boundary recall adhering the topology more. Still, the superpixels from GS04 result into very irregular shaped

regions, for which the total number of superpixels can not be defined beforehand.

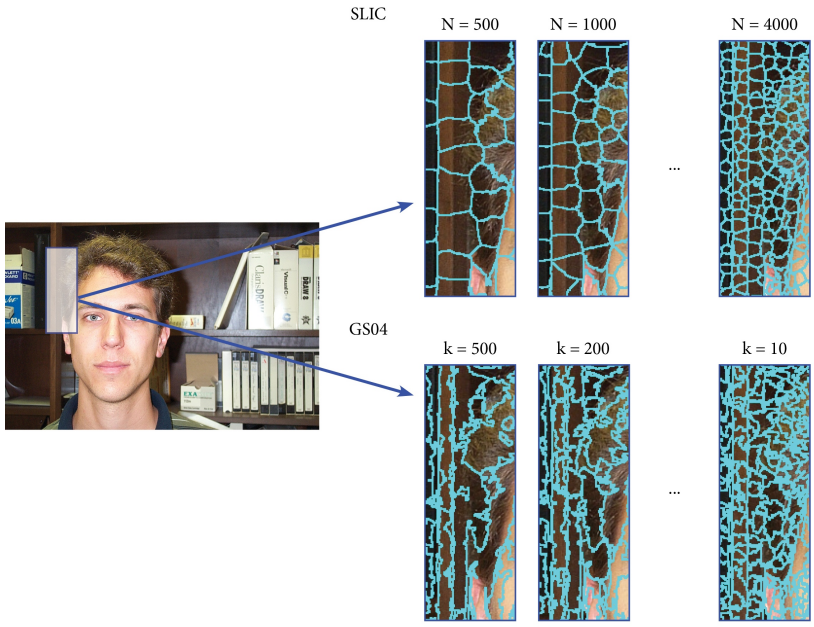


Figure 4.6: Example of granularity of superpixels and problem arising when superpixels are too big merging background and foreground into one superpixel.

CHAPTER 5

Hair and Shoulder

Hair is an important feature of human appearance and the most variant aspect of human appearance [65]. Robust and accurate hair segmentation is difficult because of challenging variation of hair color, shape [63]. Further difficulties stem from the need of a higher resolution than normally available for perfect segmentation.

In computer vision, hair style analysis or hair segmentation remains an ongoing research issue and by far an unsolved problem [63]. As described in previous chapters in literature most of the publication on skin and hair detection ignore the shoulder parts of the person, labeling it as background. In this master thesis we are looking into a full automatic *human-head* and *shoulder* segmentation and therefore propose an approach to handle hair and a correct shoulder segmentation as well.

5.1 Technical Specification

Hair has a variety of properties and attributes, which makes the detection and segmentation challenging. Yacoob and Davis [65] described hair representation along the following dimensions: length, volume, surface area, dominant color, coloring (e.g.: color variations), forehead/outer hairline, density, baldness, symmetry, split location, reflectance/shine, structural alteration (e.g.: banded, layered, or braided hair), layering arrangement, texture, sideburns, and facial hair cover. Muhammad et al. [40] characterizes hair based on different hairstyles: straight, wavy, curly, kinky, braids, dreadlocks, and short. Not only the diversity of hair patterns makes hair detection and segmentation a significant challenge within images, but also the confusion between hair and similar background. It might be possible that both the environmental background and the subject's clothing share similar color or textures to hair, and thus make a separation very difficult [66].

We specify the technical terms and conditions for the input image as follows, which will be referenced in the subsequent sections of this chapter:

1. Based on the human anatomy we know that *hair starts growing from above the forehead of the person* (except for bald and semi-bald people of course which are not considered in this thesis) and the persons *shoulders being below the face adjacent to it*. In such a static frontal-view face image, if the persons hair is long, it can appear e.g. running down over the shoulders occluding parts of it as well. If it is shorter there is the possibility that hair regions might not be connected due to occlusion of the ears or person's face.

2. Besides hair color and shape challenges, the image quality plays an important role as well to achieve an acceptable hair segmentation. On the one hand a certain *resolution* of the person is essential and on the other hand the *focus and depth of field* should ideally capture the subject in focus from nose to ears and chin to crown.

3. In this thesis we are focusing on frontal-view face images, which means that the *pose* of the person's head and shoulders captured has to face towards the camera, so the rotation of the head shall be less than $+/-5$ degrees from frontal in every direction roll, pitch and yaw. This leads to the expectation that the person illustrated in the static image has visible hair around the forehead from ear to ear and his shoulders are turned to the camera.

4. When it comes to *facial hair*, specially men can have additionally beard stubble, mustache or small to large beards, covering up parts of the face/neck or shoulders. Based on human anatomy and the pose of the person towards the camera, it can be concluded that the facial hair is either a connected region surrounded by skin or at least attached to the persons face/neck and shoulders.

5. Important is maintaining the *biometric features* of the person unchanged, since the results could be used as profile pictures, which are checked often (manually) for identification.

6. The expected output should correctly extract the person's face and shoulders with a detailed and aesthetic representation of the hair. The *aesthetic appeal* plays an important role for the eye of the beholder. So in the context of this thesis an aesthetic result of the hair and shoulder structure, means the segmented hair and outline of the shoulders has to look normal in the eyes of the beholder and e.g. not every single strand of hair is important.

5.2 Hair and Shoulder Segmentation

After the initial rough segmentation by ACM and the result of the face skin silhouette detection we obtain a trimap subdividing the picture into a known background, foreground and the undefined region, remaining with parts of the background which were not removed with ACM, the person's shoulder parts and hair (Figure 5.1).
After the skin detection a closing and fill in morphological operation is performed to

reduce small noise, close borders and classify everything inside the silhouette as face/neck by filling the holes. Only the connected component in the face location is selected as foreground to remove possible skin-alike regions outside.

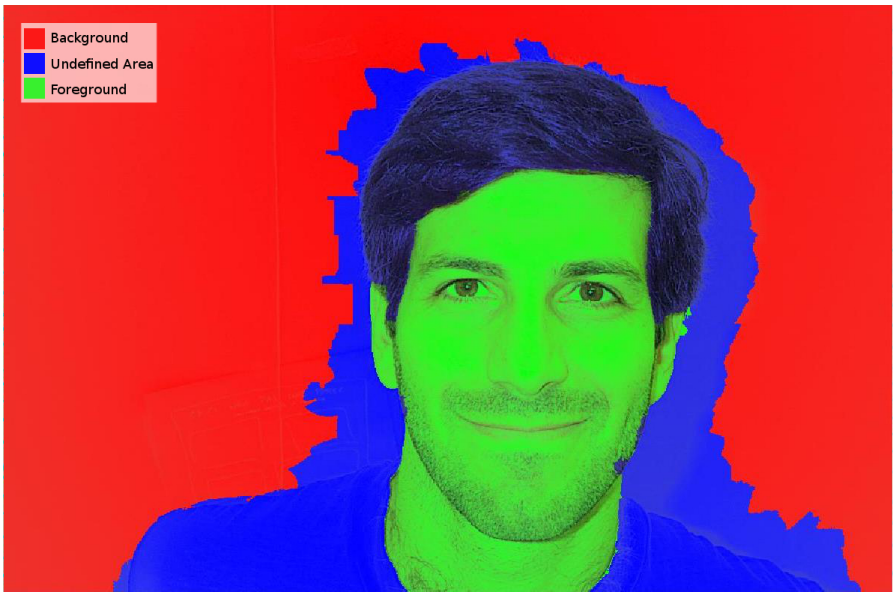


Figure 5.1: Generated trimap after ACM rough segregation and skin detection.

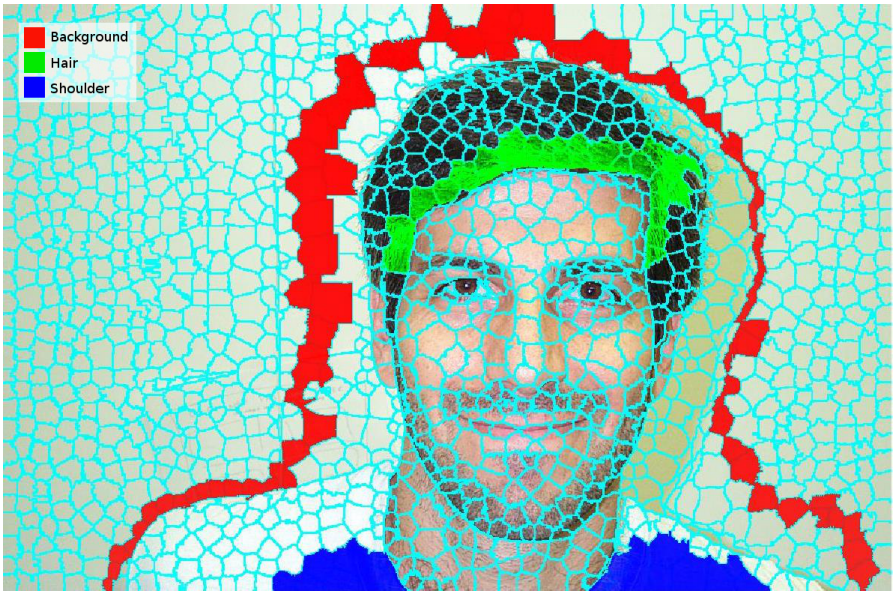


Figure 5.2: Example of the automatically selected superpixels to generate the corresponding hair, shoulder and background models to predict the rest of superpixels in the undefined area.

To classify the remaining regions of the undefined area in between the labeled background and foreground, we approximate hair being present adjacent to facial skin above the forehead to initialize a for the image particular hair model. Similarly shoulders are approximated being present below the detected face adjacent to it. This presumption follows the technical definition 1 based on the human anatomy of hair growing above the forehead and the shoulders being adjacent below the persons face. To find some reliable hair and shoulder seed pixels to obtain a generic hair and shoulder model for the particular image the results of the oversampled superpixel segmentation are considered. The initial areas are automatically set based on the location of the detected skin considering the adjacent superpixels (see Figure 5.2). Since we are focusing in this thesis on frontal-view face images the head pose plays an important role as defined in the technical specification 3. This leads to the assumption that in general the person's hair (without considering bald and semi-bald people) is visible around the forehead from ear to ear.

These *hair superpixels* or *shoulder superpixels* initialize the hair model as well as the shoulder model of the particular image. From these superpixels independently color and texture information are stored through histogram bins, to characterize the appearance of the persons hair and clothes as visualized in Figure 5.2. Additionally such a background model is computed from the surroundings of the undefined area and similarly for every *background superpixel* color and texture information is stored in histogram bins. These descriptors per superpixel per model (hair, shoulder and background) are used to identify hair, shoulder and background superpixels in the remains of the undefined area. A similarity check is performed and if the distance between a superpixel's color and texture is small, it is added to the corresponding model. Iteratively the models are refined until no changes were found. The procedure can be observed in Algorithm 5.1. In the following we are naming the superpixels from a particular model *hair*, *shoulder* or *background superpixels* and from the undefined area *undefined superpixel* respectively.

We describe *hair* and *shoulder superpixels* with color in RGB color space and texture through rotational invariant Local Binary Patterns (LBPs) [41]. Hair [46] as well as clothes [26] have particular texture aspects and it is known that the discriminating power to distinguish object classes can be improved by considering texture in addition to color [4]. It is also known that significant texture information can only be extracted if the salient object has a certain resolution and is focused in terms of depth of field (see technical specification 2).
Often in literature filter banks e.g. Leung and Malik (LM) [31] or Schmid (S) [49], which are collections of N filters, are used for texture classification. The image is convolved with multiple filters producing a stack of images for which every pixel is a feature vector of size N. Such filter banks e.g. LM consist of 48 filters, which can be time consuming. To narrow down the feature vector size and respectively the computational time, the filters with maximum response have to be evaluated for the particular problem. Observing a persons hair style in an image the hair texture can occur at arbitrary rotations and subjected to varying illumination conditions, therefore methods e.g. LM filter bank being not rotationally invariant [57] are unfavored for the texture analysis here.

Algorithm 5.1: Hair and Shoulder Detection Algorithm

input : Hair model S_{Hair}, Shoulder model $S_{Shoulder}$, Background model S_{Bg} and Undefined superpixel Set $S_{Undefined}$
output : *Undefined superpixels* classified as Hair, Shoulder and Background
parameter : Thresholds to ensure minimum dissimilarity

while $(size = |S_{Undefined}|) \geq 0$ **do**
  foreach $s_u \in S_{Undefined}$ **do**
    $s^{col}_{hair} \leftarrow$ `DissimilarityColor(`s_u, S_{Hair}, *ChiSquare*`)`;
    $s^{col}_{shoulder} \leftarrow$ `DissimilarityColor(`s_u, $S_{Shoulder}$, *ChiSquare*`)`;
    $s^{col}_{bg} \leftarrow$ `DissimilarityColor(`s_u, S_{Bg}, *ChiSquare*`)`;

    $s^{tex}_{hair} \leftarrow$ `DissimilarityTexture(`s_u, S_{Hair}, *CityBlock*`)`;
    $s^{tex}_{shoulder} \leftarrow$ `DissimilarityTexture(`s_u, $S_{Shoulder}$, *CityBlock*`)`;
    $s^{tex}_{bg} \leftarrow$ `DissimilarityTexture(`s_u, S_{Bg}, *CityBlock*`)`;

    $s^{col} \leftarrow min(s^{col}_{hair}, s^{col}_{shoulder}, s^{col}_{bg})$;
    $s^{tex} \leftarrow min(s^{tex}_{hair}, s^{tex}_{shoulder}, s^{tex}_{bg})$;

    assign s_u *to* S_{Hair}, $S_{Shoulder}$ *or* S_{Bg} *respectively if condition 5.3, 5.4, 5.5 or 5.6 holds*;
  end
  if $size == |S_{Undefined}|$ **then** *No undefined superpixel was assigned to neither Hair, Shoulder nor Background, therefore break loop and return*;
  ;
end

LBPs are a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel with the value of the center pixel and considers the result of the neighbor set as a binary number. So e.g. a LBP_8 operator produces 256 (2^8) different output values (binary patterns) with a neighborhood set of 8. When the image is rotated the intensity values will correspondingly move along the perimeter of the circle around the observed center pixel. This would result into a different LBP_8 value. To achieve rotation invariance this neighbor set is rotated clockwise so many times that a maximal number of the most significant bits are 0, which is equal to performing a circular bit-wise right shift on the binary number. For a LBP with neighborhood 8 this would lead to 36 different values, corresponding to 36 unique rotation invariant local binary patterns illustrated in Figure 5.3. These patterns can be considered also as feature detectors, e.g. #0 detects bright spots, #8 dark spots and flat areas, and #4 edges.
Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach

to the traditionally divergent statistical and structural models of texture analysis [41].

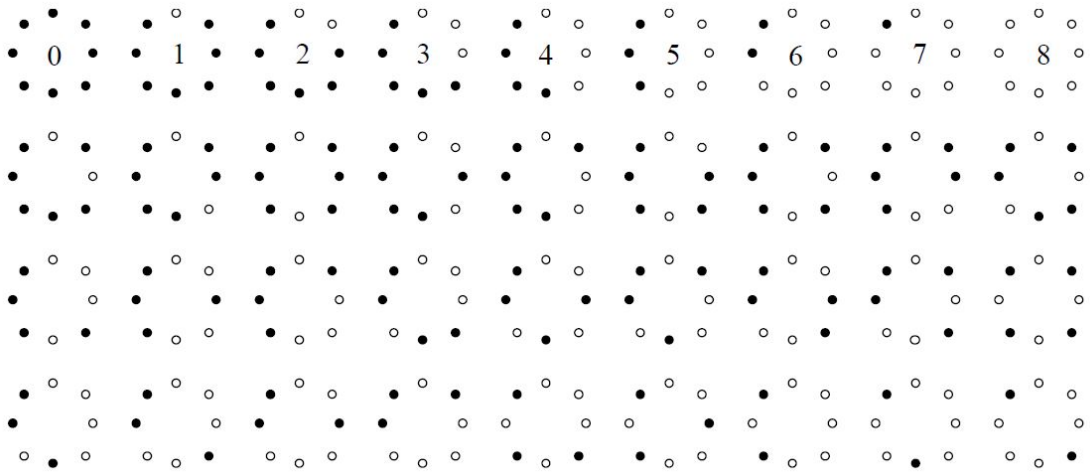


Figure 5.3: From Ojala et al. [41]: The 36 unique rotation invariant binary patterns that can occur in the eight pixel circularly symmetric neighbor set. Black and white circles correspond to bit values of 0 and 1 in the 8-bit output of the LBP_8 operator. The first row contains the nine 'uniform' patterns.

As can be observed in the description of the Algorithm 5.1 we chose two different distance metrics for measuring dissimilarity for color and texture histograms (Pseudocode lines $3-5$ *DissimilarityColor* and pseudocode lines $6-8$ *DissimilarityTexture*): For color dissimilarity Chi-Square distance was used and for texture dissimilarity City Block distance was more suiting. The Chi-squared distance between two normalized histogram samples x and y can be computed by:

$$\textit{ChiSquare:} \quad \chi^2(x,y) = \frac{1}{2}\sum_{i=1}^{d}\frac{(x_i - y_i)^2}{x_i + y_i} \tag{5.1}$$

where i denotes a d-dimensional vector, which in our case is the number of histogram bins, and x_i indicates the i-th feature of the sample x. Chi-Square distance metric is bounded to $[0;1]$, 0 meaning both histogram samples are equal showing no dissimilarity and 1 meaning both histograms are completely different not intersecting at all.
City block distance, in literature also referred to as Manhattan distance, between two normalized histogram samples x and y can be computed as follows:

$$\textit{CityBlock:} \quad D(x,y) = \sum_{i=1}^{d}|x_i - y_i| \tag{5.2}$$

where D is always greater than or equal to zero. The measurement would be 0 for identical samples and high for samples that show little similarity. In most cases, this distance measure yields results similar to the Euclidean distance (also referred in literature as L2

distance). However, note, that with City block distance, the effect of a large difference in a single dimension is dampened, since the distances are not squared.
For illustration purposes we simplified a possible color histogram dissimilarity check (see histogram samples in Figure 5.4), for which the Chi-Square χ^2, City Block D and Euclidean Distance $L2$ between samples x and y_1, y_2 are computed, to demonstrate the importance of choosing the best suiting distance metric. Per superpixel a color feature vector is described by its color histogram bins and generally since SLIC as well as GS04 are oversampling the image using color information, the histogram would look similar to sample x in Figure 5.4. Observing the three histogram samples we would suggest that the color histogram x is more similar to y_1 than y_2, but only for the Chi-Square distance this is true.
Regarding the texture dissimilarity check, City Block distance metric is more suiting since the texture histogram is generally more uniformly distributed.

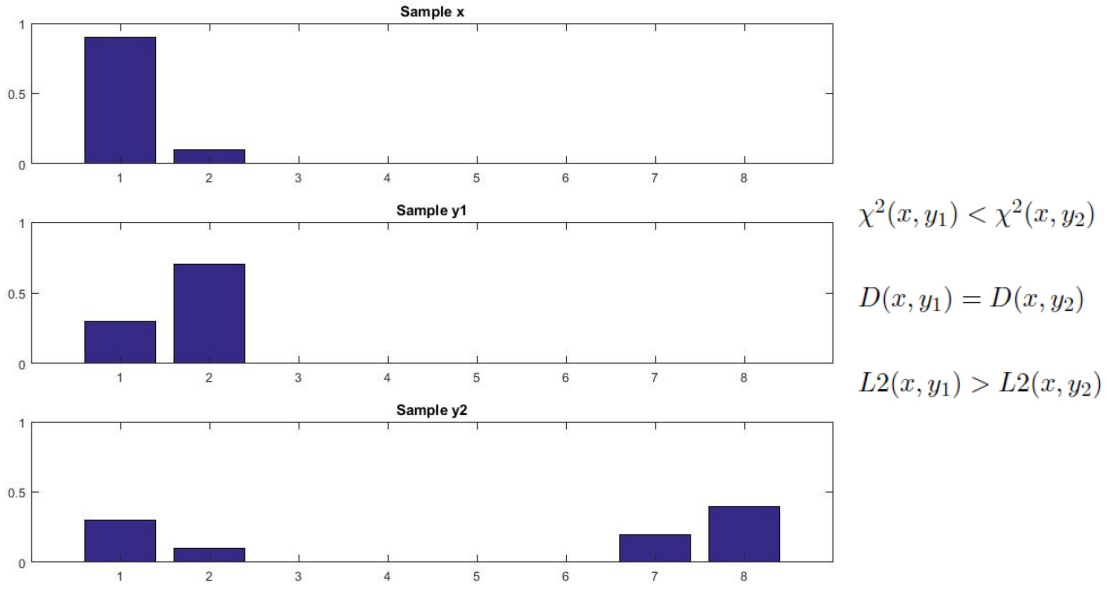


Figure 5.4: Simplified example color histograms with 8 bins and their distance metric results comparing Chi-Square χ^2, City Block D and Euclidean Distance $L2$.

Some logic constraints based on the human anatomy were defined, such as our initial assumptions of hair growing from above the forehead of the person and his shoulders being adjacent below to the face (see technical specification 1). Additionally, we split the image into a *hair region* and *shoulder region* not allowing a superpixel in the undefined *hair region* to match in similarity with a shoulder superpixel from the *shoulder region*. The other way around is possible since the person's hair can reach down to the shoulders or even occlude parts of them.
A *undefined superpixel* is labeled hair, shoulder or background if the following respective condition holds:

$$\begin{aligned}(s_u^{tex} \leq \theta_h^{tex}) \quad &\wedge \quad (min(s^{tex}) \in S_{Hair}) \quad \wedge \\ (s_u^{col} \leq \theta_h^{col}) \quad &\wedge \quad (min(s^{col}) \in S_{Hair}) \\ &\Rightarrow s_u \in S_{Hair}\end{aligned} \tag{5.3}$$

$$\begin{aligned}(s_u^{tex} \leq \theta_s^{tex}) \quad &\wedge \quad (min(s^{tex}) \in S_{Shoulder}) \quad \wedge \\ (s_u^{col} \leq \theta_s^{col}) \quad &\wedge \quad (min(s^{col}) \in S_{Shoulder}) \quad \wedge \\ &(s_u \in \text{ShoulderRegion}) \\ &\Rightarrow s_u \in S_{Shoulder}\end{aligned} \tag{5.4}$$

$$\begin{aligned}(s_u^{tex} \leq \theta_h^{tex}) \quad &\wedge \quad (min(s^{tex}) \in S_{Bg}) \quad \wedge \\ (s_u^{col} \leq \theta_h^{col}) \quad &\wedge \quad (min(s^{col}) \in S_{Bg}) \\ &\Rightarrow s_u \in S_{Bg}\end{aligned} \tag{5.5}$$

where s_u is a *undefined superpixel* and S_{Hair}, $S_{Shoulder}$ and S_{Bg} are being the sets of hair, shoulder and background superpixels. For the first constraint in Formula 5.3 a *undefined superpixel* is added to the set of hair superpixels S_{Hair} if the texture dissimilarity s_u^{tex} is under a certain threshold θ_h^{tex}, the minimal texture dissimilarity is with a superpixel s^{tex} from set S_{Hair} and for the color dissimilarity of the undefined superpixel s_u^{col} the same has to hold. A *undefined superpixel* s_u is added to the shoulder set $S_{Shoulder}$ if additionally to the conditions above s_u is in the defined shoulder region of the image, otherwise it cannot be part of the person's shoulder since this is from anatomy perspective of the human body not possible (see technical specification 1).
It can be observed that we are looking at the color and texture characteristics independently. It is possible that $min(s^{tex})$ and $min(s^{col})$ are different superpixels. Otherwise we would need to set a relation between texture and color when combining the two results together. The question would arise how to weight them? How can both normalized values be compared when they are so independent from each other. We have evaluated this during our master thesis and came to the conclusion that this is not resulting into a possibility to assign hair, shoulder and background superpixels correctly. The problem arising is that if both are combined it might occur frequently that a supposedly hair superpixel would be labeled as background, even though e.g. the color similarity is not minimal.
The thresholds θ_h^{tex}, θ_h^{col} for possible hair similarity and respectively θ_s^{tex}, θ_s^{col} for shoulder similarity ensure that the *undefined superpixel* has to have a certain similarity to the foreground or background superpixel. Also the fact that it has to be labeled twice for color and texture with the corresponding region ensures a stability of the procedure.
A dissimilarity in e.g. color of $\theta_h^{col} = 0.4$ means that the distance function of a unlabeled superpixel and its corresponding match has to be similar for at least 60%. A larger

dissimilarity threshold would work well for images with simpler (uniform) color regions leading to fewer iterations of the Algorithm 5.1 to generate a result. For more complex problem regions as it is for hair, clothes and complex backgrounds as we are phasing, a smaller threshold should be used. Specially, when it comes to low contrast around the silhouette of the subject (e.g. background has a similar color compared to the person's hair) then the normalized differences are reduced, which leads to a higher acceptance rate of supposedly background superpixels if the threshold is to large.
Important to consider here is the size of the superpixels as well. The smaller the superpixels the larger the dissimilarity thresholds should be, because for bigger superpixels the similarity in general is higher.

Hair and clothes have different characteristics when it comes to color and texture. Normally hair is higher frequential. The shoulders, depending on the material of texture of the clothes the person is wearing, have in general the highest frequencies around the border if the contrast to the background is high. Therefore one more constraint concerning a undefined superpixel s_u located in the *ShoulderRegion* prioritizing color is added as forth condition during the dissimilarity test:

$$\begin{aligned}(s_u^{col} \leq \frac{\theta_s^{col}}{2}) \quad \wedge \quad (min(s^{col}) \in S_{Shoulder}) \quad \wedge \quad (s_u \in \text{ShoulderRegion}) \\ \Rightarrow s_u \in S_{Shoulder}\end{aligned} \tag{5.6}$$

where only color information is considered if the *undefined superpixel* color dissimilarity s_u^{col} is under half of θ_s^{col} and the minimum dissimilarity matches a superpixel in the shoulder set $S_{Shoulder}$. If this condition holds then s_u is added to the $S_{Shoulder}$ as well, regardless of the texture information. The low threshold of $\frac{\theta_s^{col}}{2}$ ensures that only very similar in color superpixels are added to the $S_{Shoulder}$ set.

After every iteration *undefined superpixels* are assigned to a either foreground (hair or shoulder set) or background and the algorithm terminates if no changes were made. For all still unassigned undefined superpixels a probabilistic mask is computed independently for color and texture since they were assigned to different classes or the similarity was above the thresholds.
This probabilistic mask is considered in the post-processing step for minor improvements on the result by additionally considering the location of the particular superpixels and its neighborhood. If the superpixel is e.g. surrounded by hair and it has a high probability to be actually hair, either because of the color or texture similarity, then it is assigned to hair, otherwise to the background (see Figure 5.5).

(a) Original Image.

(b) The higher the intensity, the higher the likelihood.

(c) Human-head and shoulder segmentation result.

Figure 5.5: Example of the probabilistic mask (b) and the result (c) of our procedure on a particular sample image from CALTECH database

5.3 Results and Evaluation

In the following subsections we describe the dataset of frontal-view face images used for the quantitative and qualitative results. We evaluate the proposed method by assessing the consistency of segmentation results and the manually labeled ground truth in terms of the overlap ratio, which was used in this context before by Xin et al. [64]. This evaluation criterion measures the error when segmented foreground contains background information and vice versa:

$$overlap = \frac{Ground \cap Segment}{Ground \cup Segment} \tag{5.7}$$

where *Ground* is the image ground truth, and *Segment* is our segmentation result. In our case well known evaluation measurements e.g. accuracy, precision, recall or F1 score (harmonic average of the precision and recall) would not return meaningful information about the performance of the approach, since in all cases it would approximate 1. This is because most of the image pixels are labeled correctly as background or foreground and the only small erroneous area is around the silhouette of the person, that is why the overlap ratio was chosen as a significant evaluation criterion. Based on this criterion and the qualitative results we compare our methodology using both discussed superpixel algorithms SLIC and GS04 (see Chapter 4).

5.3.1 Database

Experiments are conducted using a public dataset for which ground truth was manually segmented for 250 images totally. Representative sample images were selected to demonstrate the performance and limitations of our approach reaching to a variation of ethnicity, gender, age, illumination, (simple) to complex backgrounds, outdoor and indoor scenes, long/short hair with different hairstyles, color and facial hair (three-days-beard, mustache). The database was used previously in the skin detection evaluation (see Chapter 3) and in the comparison of the superpixel algorithms in Chapter 4 as well.

- *CALTECH* [1]*:* is a frontal face dataset collected at California Institute of Technology, capturing 27 people under different light conditions, facial expression, ethnicity groups (mostly white and Asian), gender and complex backgrounds. It provides images under different conditions with a complex background, where the orientation of the head and shoulders is facing the camera according to the defined criteria we are focusing on in this thesis. The database does not provide any ground truth. Therefore, ground truth was generated for a set of 50 images. Additionally, to enlarge the dataset for significant quantitative and qualitative results, the background is replaced with arbitrary chosen complex background images from the Berkeley Benchmark [38] resulting into a total number of 250 images (see Figure 5.6).

[1] Collected by Markus Weber at California Institute of Technology `http://www.vision.caltech.edu/html-files/archive.html`

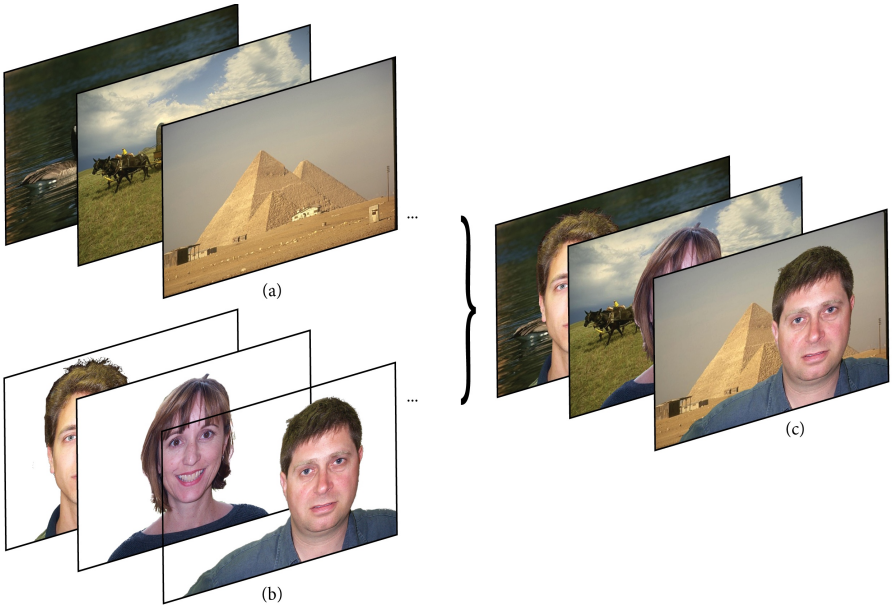


Figure 5.6: Enlarge dataset (a) combining complex backgrounds from Berkeley Benchmark Dataset and (b) the manually segmented ground truth from CALTECH (c) resulting into new images for quantitative and qualitative evaluations.

5.3.2 Human-head and Shoulder Segmentation Results

For the following quantitative and qualitative results the decision tree was used as classification learner for the skin detection. Furthermore, concluding from the results in Chapter 4 for the superpixel algorithm SLIC $N = 1500$ of total superpixels was selected and for GS04 $k = 50$. The dissimilarity threshold values were chosen by experience: $\theta_h^{tex} = 0.35$, $\theta_h^{col} = 0.4$, $\theta_s^{tex} = 0.5$, $\theta_s^{col} = 0.4$.

Experiments are carried out on 250 frontal-view face images comparing the segmentation result with the manually generated ground truth. As described above the overlap ratio was considered as evaluation criterion and the results are visualized in Figure 5.7. As can be observed for this particular dataset the average ratio for the head and shoulder segmentation algorithm using GS04 with 0.9482 outperforms using SLIC with 0.9449 by a very small margin. To give the reader an impression on how such results look like with a high and low overlap ratio, the best as well as the worst results are illustrated. The skin alike colors in the background of the worst case sample lead to a failed skin detection result, from which in the following no hair samples can be found to initialize the hair model correctly. Observing the best sample result a almost perfect segmentation is achieved.

Figure 5.8 shows some qualitative results comparing our algorithm using both superpixel algorithms. It can be observed that indoor as well as outdoor scenes achieve good results. In most of the indoor scenes in this particular dataset the subject casts a shadow due

to the used illumination conditions (see Figure 5.8 first, second and seventh sample). The color dissimilarity of these regions with rest of the background is often high and could be falsely detected as more similar to the hair, but with the additional texture dissimilarity constraint such problems can be prevented easily as can be seen in e.g. the second row. In sample six and eight we illustrate two results of persons with longer hair, where specially observing the man in the last row his hair is detected even though the hair regions are not connected to each other, nevertheless all hair regions are detected and labeled as foreground.
Small erroneous regions are visible e.g. in the fifth sample only for the result of the segmentation using SLIC superpixel algorithm, where parts of the background on the left side are detected as part of the shoulders. Similar is the incorrect labeling of a superpixel above the head of the seventh sample.

A possible problem occurs when the models are initialize incorrectly containing superpixels which are falsely considered in the model as e.g. in Figure 5.8 last sample for the segmentation using SLIC to generate the superpixels. In this particular sample this is not the case for the segmentation using GS04. Similar is the behavior described above, if the skin detection fails completely and no hair superpixel is added to the hair model (see worst sample in Figure 4.4). In Figure 5.9 the initialization mask for the ACM resulting into the rough segmentation as one of the preprocessing steps, defines partly strands of hair as background. The superpixels generated with SLIC containing these strands of hair bias the segmentation in an non-acceptable way. As for the segmentation with GS04, even though the background model is initialized wrongly with hair superpixels, it does not affect the segmentation much, compensating the error in an acceptable result.

5.4 Discussion

With our proposed algorithm to segment hair and shoulders in a static frontal-view face image without prior knowledge on the person's appearance and background complexity, we are able to handle all sorts of different input images following the conditions stated in the technical specification Section 5.1. We evaluated our proposed algorithm with the overlap ratio in a quantitative way and showed representative qualitative result.
Still, we have to keep in mind that this evaluation is limited in a way since the possibilities of input images are endless. Therefore, the logical constraints and properties of our proposed hair and shoulder algorithm can be taken into consideration. Based on the properties of our algorithm we can conclude that hats or other head coverings could also be handled if they are not casting significant shadows in the face region, because then instead of *hair superpixels* initializing the hair model, *hat or head covering superpixels* describe the model. Similarly would be the initialization of the shoulder model of an image with a bearded person (see technical specification 4). After the face skin detection the shoulder model would be initialized partly with superpixels containing information of the person's beard and partly the person's clothing. Following our algorithm iteratively would lead to finding the remaining *beard superpixels* as well as the *shoulder superpixels*.

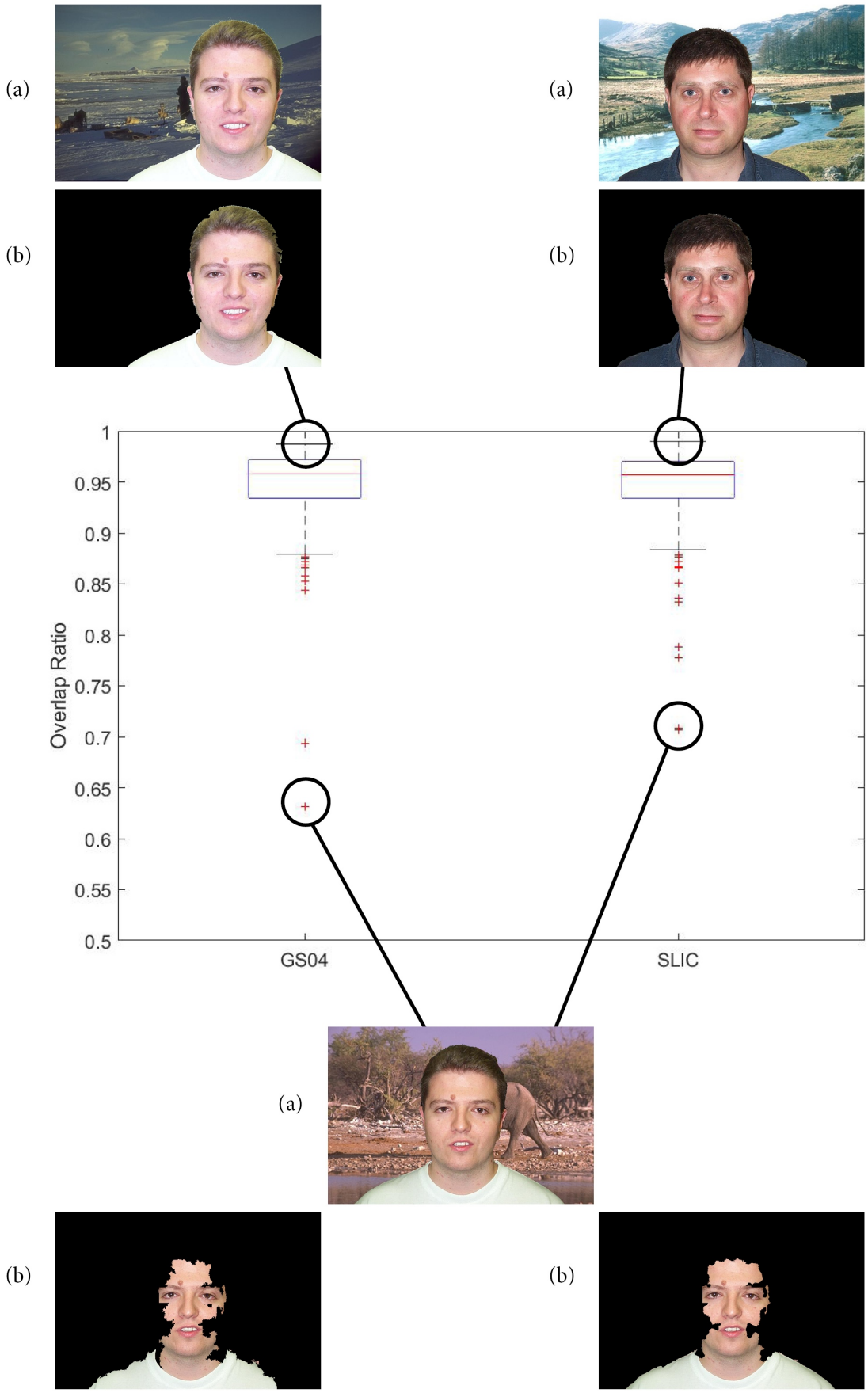
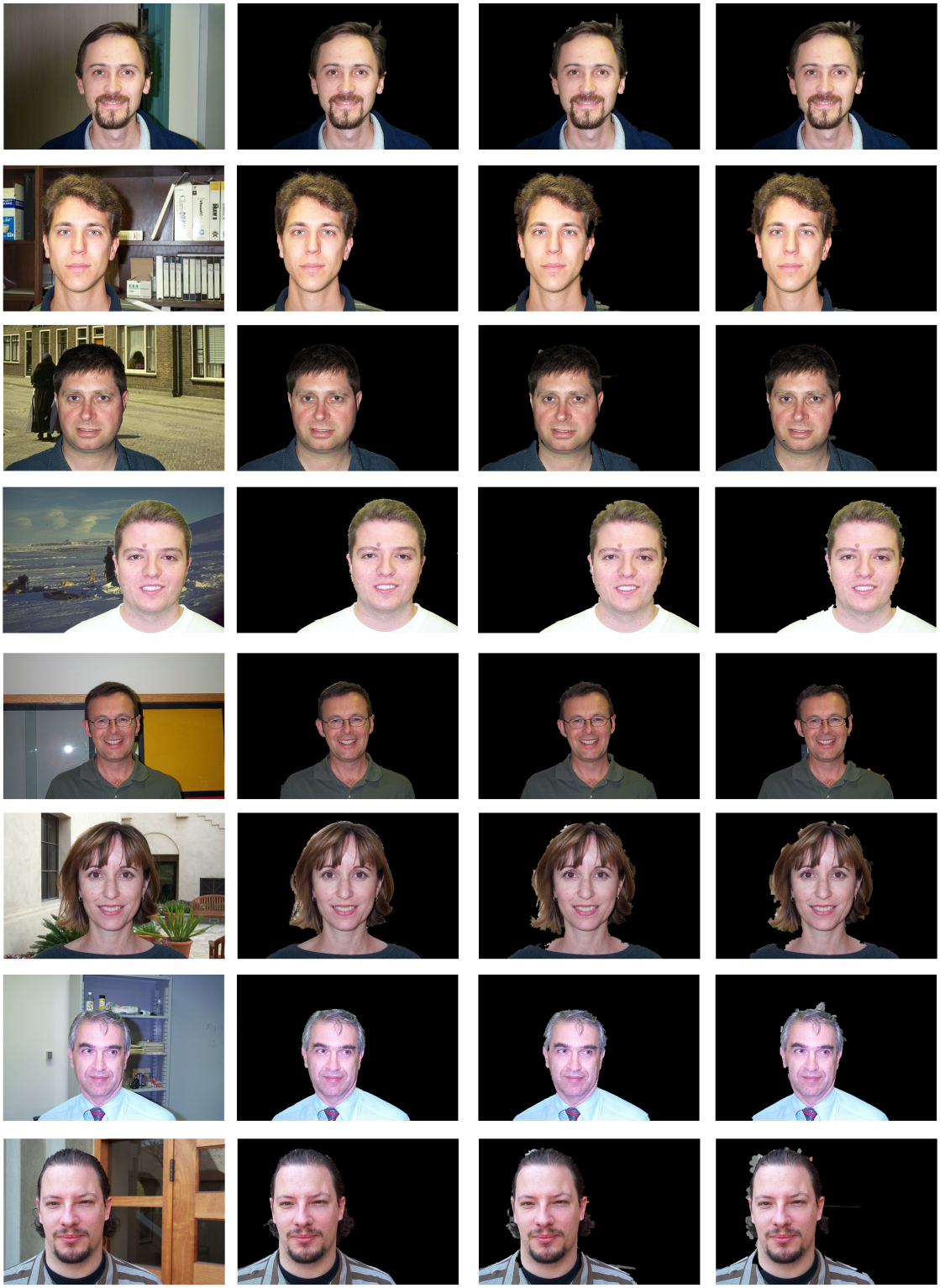


Figure 5.7: Overlap Ratio of our Human-Head and Shoulder Segmentation using GS04 and SLIC superpixel algorithm. Best and worst overlap ratio result is above and below the boxplot respectively with (a) Original Image and (b) the segmentation result.



(a) (b) (c) (d)

Figure 5.8: Head and shoulder segmentation results.(a) Original image. (b) Ground truth. (c) using GS04 (d) with SLIC as oversampling algorithm.

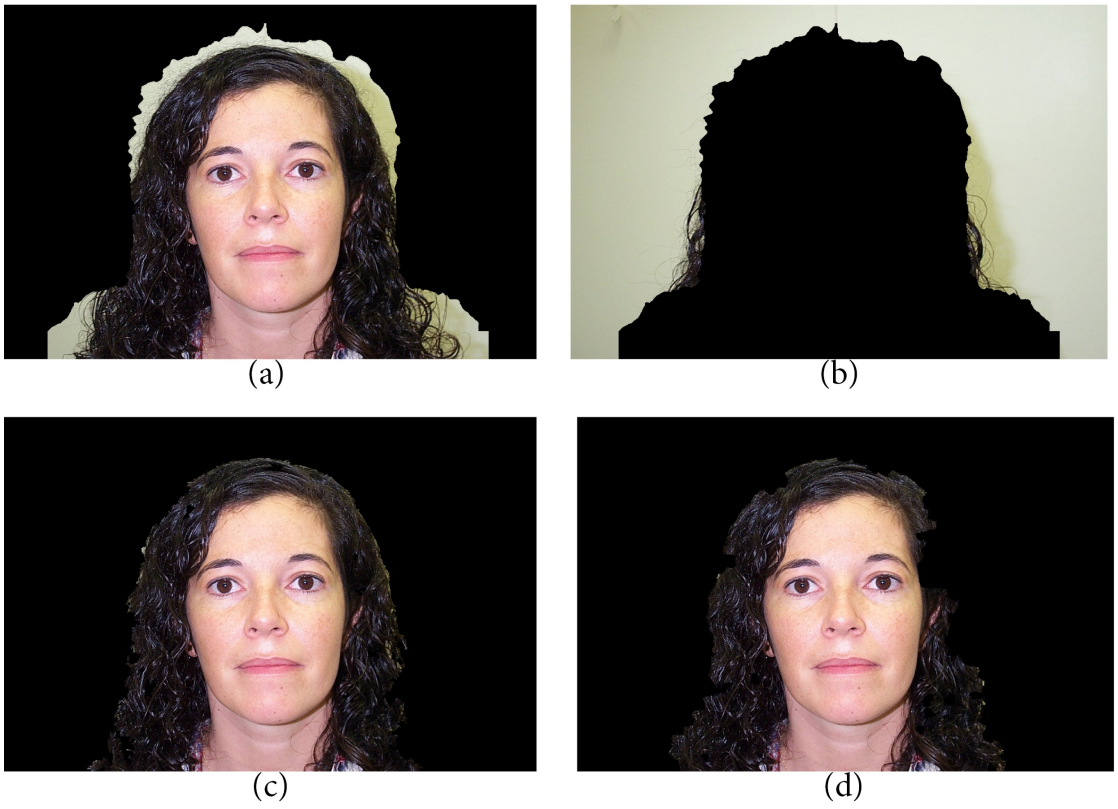


Figure 5.9: After incorrect rough segregation with ACM due to narrow initialization mask (a) ACM result *foreground mask* (containing the person and possible remaining background). (b) ACM result *background mask* (containing only background regions). (c) Head and shoulder segmentation with GS04 superpixel algorithm. (d) Head and shoulder segmentation with SLIC superpixel algorithm.

CHAPTER 6

Conclusion

In this thesis, we present a novel automatic human-head and shoulder segmentation algorithm of frontal-view portrait images with arbitrary unknown complex backgrounds. We began by introducing the basic strategy of our approach and how the different subtasks are combined to form the proposed methodology for the expected result. We discussed these subtasks of our method independently in the main section of the thesis, as each task can be viewed as a subtask of our particular problem, but also as a possible approach for similar technical requirements: Face Skin Silhouette Detection in Chapter 3; oversampling the input image and comparing two different state-of-the-art superpixel algorithms in Chapter 4; and Hair and Shoulder Segmentation in Chapter 5. We evaluate and provide qualitative and quantitative results using different public databases for each individual part to illustrate their performance.

As one main contribution, we introduce a new Face Skin Silhouette Detection algorithm based on supervised classification learners (FSCL), adding automatically labeled pixel information from the query image to the training set, improving the performance of the classifiers prediction on the remaining query image significantly. Another contribution is the comparison and discussion of two different superpixel algorithms, SLIC and GS04, on their adherence of boundaries, evaluating the boundary recall and under-segmentation error in general and the overlap ratio for our particular problem statement highlighted in this thesis. A further major contribution is our hair, shoulder and background models in our novel Hair and Shoulder Segmentation algorithm, composed of color, texture and the superpixel's relative position in the query image, without any prior knowledge on the person and background complexity. With these image specific models, the remaining of the query image is classified into foreground and background. Based on the logical conditions and algorithm properties, an additional feature of our approach is that due to occlusion not connected regions can be found correctly and labeled with the same class. We achieve an automatic human-head and shoulder segmentation without changing

any biometric features on the persons face to allow a possible identification in the following.

Regarding future work we would like to address the problem of handling baldness and semi-baldness as part of our human-head and shoulder segmentation. Moreover, incorporating alpha matting or guided image filtering [17] to create a smooth transition between foreground and any new background may greatly improve the visual quality of the results.

We want to conclude this thesis by broadening the context. Observing our Hair and Shoulder Segmentation algorithm, the combination of color and texture information extracted from superpixels to initialize e.g. the hair model, is able to label the remaining undefined unlabeled areas of the image as hair, even though parts of those segments do not have to be connected. This could be useful for other applications which e.g. have a salient object which is partly occluded, resulting in multiple components. Furthermore in our case the models are initialized only with information on the particular query image but they can be initialized easily with any additional training data depending on the application. Logical conditions can be added easily as well as their weight can be modified as we performed different threshold parameter settings for classifying clothes for the shoulder part or hair.

List of Figures

List of Tables

List of Algorithms

Bibliography

[1] P. Aarabi. Automatic segmentation of hair in images. In *2015 IEEE International Symposium on Multimedia (ISM)*, pages 69–72, Dec. 2015.

[2] M. Abdullah-Al-Wadud, M. Shoyaib, and O. Chae. A skin detection approach based on color distance map. *EURASIP J. Adv. Signal Process*, 2008:199:1–199:10, Jan. 2008. ISSN 1110-8657.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, Nov. 2012.

[4] I. Ahn and C. Kim. Face and hair region labeling using semi-supervised spectral clustering-based multiple segmentations. *IEEE Transactions on Multimedia*, 18(7): 1414–1421, 2016.

[5] M. Barstugan, R. Ceylan, M. Sivri, and H. Erdogan. Automatic liver segmentation in abdomen ct images using slic and adaboost algorithms. In *Proceedings of the 2018 8th International Conference on Bioscience, Biochemistry and Bioinformatics*, ICBBB 2018, pages 129–133, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5341-0.

[6] R. B. Bhatt, G. Sharma, A. Dhall, and S. Chaudhury. Efficient Skin Region Segmentation Using Low Complexity Fuzzy Decision Tree Model. In *2009 Annual IEEE India Conference*, pages 1–4, Dec. 2009.

[7] P. Bu, N. Wang, and H. Ai. Using Structural Patches Tiling to Guide Human Head-shoulder Segmentation. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 797–800, New York, NY, USA, 2012. ACM.

[8] M. Chai, T. Shao, H. Wu, Y. Weng, and K. Zhou. Autohair: fully automatic hair modeling from a single image. *ACM Transactions on Graphics (ToG)*, 35(4):116, 2016.

[9] J. Chatrath, P. Gupta, P. Ahuja, A. Goel, and S. M. Arora. Real time human face detection and tracking. In *2014 International Conference on Signal Processing and Integrated Networks (SPIN)*, pages 705–710, Feb. 2014.

[10] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):603–619, May 2002. ISSN 0162-8828.

[11] X. Deng and X. Wu. Fast Head-and-shoulder Segmentation. Master's thesis, McMaster University, Canada, 2016.

[12] A. Diplaros, T. Gevers, and N. Vlassis. Skin detection using the EM algorithm with spatial constraints. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, volume 4, pages 3071–3075 vol.4, Oct. 2004.

[13] A. Elgammal, C. Muang, and D. Hu. Skin detection-a short tutorial. *Encyclopedia of Biometrics*, pages 1–10, 2009.

[14] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, Sept. 2004. ISSN 0920-5691.

[15] F. Gasparini and R. Schettini. Skin segmentation using multiple thresholding. In *Internet Imaging VII, Proceedings of SPIE*, volume 6061, pages 128–135, 2006.

[16] H. Greenspan, J. Goldberger, and I. Eshet. Mixture model for face-color modeling and segmentation. *Pattern Recognition Letters*, 22(14):1525–1536, 2001.

[17] K. He, J. Sun, and X. Tang. Guided image filtering. In *European conference on computer vision*, pages 1–14. Springer, 2010.

[18] International Civil Aviation Organization (ICAO). *Machine Readable Travel Documents: Part 5.* ICAO, 999 Robert-Bourassa Boulevard, Montréal, Quebec, Canada H3C 5H7, 7 edition, 2015. Doc 9303.

[19] ISO/IEC 19794-5. *Information technology - Biometric data interchange formats - Part 5: Face image data.* ISO, 1 edition, June 2005.

[20] J. C. S. Jacques and S. R. Musse. Improved Head-Shoulder Human Contour Estimation through Clusters of Learned Shape Models. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pages 329–336, Aug. 2015.

[21] J. C. S. Jacques, C. R. Jung, and S. R. Müsse. Head-shoulder human contour estimation in still images. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 278–282, Oct. 2014.

[22] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning: with Applications in R.* Springer, New York, 1st ed. 2013, corr. 7th printing 2017 edition, Sept. 2017. ISBN 978-1-4614-7137-0.

[23] D. Jensch, D. Mohr, and G. Zachmann. A Comparative Evaluation of Three Skin Color Detection Approaches. *Journal of Virtual Reality and Broadcasting*, 12(2015) (1), Jan. 2015.

[24] P. Julian, C. Dehais, F. Lauze, V. Charvillat, A. Bartoli, and A. Choukroun. Automatic hair detection in the wild. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 4617–4620. IEEE, 2010.

[25] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A Survey of Skin-color Modeling and Detection Methods. *Pattern Recogn.*, 40(3):1106–1122, Mar. 2007.

[26] Y. Kalantidis, L. Kennedy, and L.-J. Li. Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos. In *Proceedings of the 3rd ACM conference on International conference on multimedia retrieval*, pages 105–112. ACM, 2013.

[27] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, Jan. 1988.

[28] R. Khan, A. Hanbury, and J. Stöttinger. Universal seed skin segmentation. *Advances in Visual Computing*, pages 75–84, 2010.

[29] R. Khan, A. Hanbury, and J. Stöttinger. Skin detection: A random forest approach. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 4613–4616. IEEE, 2010.

[30] J. Y. Lee and S. I. Yoo. An elliptical boundary model for skin color detection. In *In Proc. Int. Conf. on Imaging Science, System and Technology*, 2002.

[31] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International journal of computer vision*, 43(1):29–44, 2001.

[32] A. Levinshtein, C. Chang, E. Phung, I. Kezele, W. Guo, and P. Aarabi. Real-time deep hair matting on mobile devices. *CoRR*, 2017.

[33] L. Liang, C. F. Huitema, M. A. Simari, and S. E. Anderson. Camera system and method for hair segmentation, Sept. 19 2017. US Patent 9,767,586.

[34] C. Liensberger, J. Stöttinger, and M. Kampel. Color-based and context-aware skin detection for online video annotation. In *2009 IEEE International Workshop on Multimedia Signal Processing*, pages 1–6, Oct. 2009.

[35] W. Lü and J. Huang. Skin detection method based on cascaded adaboost classifier. *Journal of Shanghai Jiaotong University (Science)*, 17(2):197–202, Apr 2012.

[36] B. Ma, C. Zhang, J. Chen, R. Qu, J. Xiao, and X. Cao. Human skin detection via semantic constraint. In *Proceedings of International Conference on Internet Multimedia Computing and Service*, ICIMCS '14, pages 181:181–181:184, New York, NY, USA, 2014. ACM.

[37] I. Maglogiannis. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. Frontiers in artificial intelligence and applications. IOS Press, 2007.

[38] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.

[39] R. Melán. Skin detection in frontal-view faces. Technical Report PRIP-TR-142, PRIP, TU Wien, 2018.

[40] U. R. Muhammad, M. Svanera, R. Leonardi, and S. Benini. Hair detection, segmentation, and hairstyle classification in the wild. *Image and Vision Computing*, 71:25 – 37, 2018.

[41] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *European Conference on Computer Vision*, pages 404–420. Springer, 2000.

[42] P. O. Pinheiro, R. Collobert, and P. Dollár. Learning to segment object candidates. In *Advances in Neural Information Processing Systems*, pages 1990–1998, 2015.

[43] P. O. Pinheiro, T.-Y. Lin, R. Collobert, and P. Dollár. Learning to refine object segments. In *European Conference on Computer Vision*, pages 75–91. Springer, 2016.

[44] C. Platzer, M. Stuetz, and M. Lindorfer. Skin Sheriff: A Machine Learning Solution for Detecting Explicit Images. In *Proceedings of the 2Nd International Workshop on Security and Forensics in Communication Systems*, SFCS '14, pages 45–56, New York, NY, USA, 2014. ACM.

[45] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 309–314, New York, NY, USA, 2004. ACM.

[46] C. Rousset and P. Y. Coulon. Frequential and color analysis for hair mask segmentation. In *2008 15th IEEE International Conference on Image Processing*, pages 2276–2279, Oct 2008.

[47] A. A. Sangüesa, N. K. Jorgensen, C. A. Larsen, K. Nasrollahi, and T. B. Moeslund. Initiating grabcut by color difference for automatic foreground extraction of passport imagery. In *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6, Dec 2016.

[48] F. Saxen and A. Al-Hamadi. Color-based skin segmentation: An evaluation of the state of the art. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 4467–4471, Oct. 2014.

[49] C. Schmid. Constructing models for content-based image retrieval. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–II. IEEE, 2001.

[50] K. B. Shaik, P. Ganesan, V. Kalist, B. S. Sathish, and J. M. M. Jenitha. Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space. *Procedia Computer Science*, 57:41–48, Jan. 2015.

[51] Y. Shen, Z. Peng, and Y. Zhang. Image based hair segmentation algorithm for the application of automatic facial caricature synthesis. *The Scientific World Journal*, 2014, 2014.

[52] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, Aug. 2000.

[53] G. Shu. *Human Detection, Tracking and Segmentation in Surveillance Video*. PhD thesis, M.S. Shanghai Jiaotong University, China, 2014.

[54] L. Spillmann and J. S. Werner. *Visual Perception: The Neurophysiological Foundations*. Academic Press, 2 edition, 1990.

[55] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell. A Fusion Approach for Efficient Human Skin Detection. *IEEE Transactions on Industrial Informatics*, 8(1): 138–147, Feb. 2012.

[56] D. V. Thombre, J. H. Nirmal, and D. Lekha. Human detection and tracking using image segmentation and Kalman filter. In *2009 International Conference on Intelligent Agent Multi-Agent Systems*, pages 1–5, July 2009.

[57] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International journal of computer vision*, 62(1-2):61–81, 2005.

[58] A. Vedaldi and S. Soatto. *Quick Shift and Kernel Methods for Mode Seeking*, pages 705–718. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[59] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13(6):583–598, June 1991.

[60] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.

[61] N. Vu and B. S. Manjunath. Shape prior segmentation of multiple objects with graph cuts. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, June 2008.

[62] D. Wang, S. Shan, W. Zeng, H. Zhang, and X. Chen. A novel two-tier bayesian based method for hair segmentation. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 2401–2404. IEEE, 2009.

[63] N. Wang, H. Ai, and S. Lao. A compositional exemplar-based model for hair segmentation. In *Asian Conference on Computer Vision*, pages 171–184. Springer, 2010.

[64] H. Xin, H. Ai, H. Chao, and D. Tretter. Human head-shoulder segmentation. In *Face and Gesture 2011*, pages 227–232, Mar. 2011.

[65] Y. Yacoob and L. S. Davis. Detection and analysis of hair. *IEEE transactions on pattern analysis and machine intelligence*, 28(7):1164–1169, 2006.

[66] C.-K. Yang and C.-N. Kuo. Automatic hair extraction from 2d images. *Multimedia Tools and Applications*, 75(8):4441–4465, 2016.