SKIN DETECTION IN IMAGE AND VIDEO FOUNDED IN
CLUSTERING AND REGION GROWING

ABM Rezbaul Islam

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2019

APPROVED:

Bill Buckles, Major Professor

Armin R. Mikler, Committee Member

Robert Akl, Committee Member

Kamesh Namuduri, Committee Member

Barrett Bryant, Chair of the Department
of Computer Science and
Engineering

Hanchen Huang, Dean of the College of
Engineering

Victor Prybutok, Dean of the Toulouse
Graduate School

Islam, ABM Rezbaul. *Skin Detection in Image and Video Founded in Clustering and Region Growing*. Doctor of Philosophy (Computer Science and Engineering), August 2019, 107 pp., 18 tables, 54 figures, 97 numbered references.

Researchers have been involved for decades in search of an efficient skin detection method. Yet current methods have not overcome the major limitations. To overcome these limitations, in this dissertation, a clustering and region growing based skin detection method is proposed. These methods together with a significant insight result in a more effective algorithm. The insight concerns a capability to define dynamically the number of clusters in a collection of pixels organized as an image. In clustering for most problem domains, the number of clusters is fixed a priori and does not perform effectively over a wide variety of data contents. Therefore, in this dissertation, a skin detection method has been proposed using the above findings and validated. This method assigns the number of clusters based on image properties and ultimately allows freedom from manual thresholding or other manual operations. The dynamic determination of clustering outcomes allows for greater automation of skin detection when dealing with uncertain real-world conditions.

Copyright 2019
by
ABM Rezbaul Islam

ACKNOWLEDGEMENTS

First of all, I like to thank Almighty Allah for allowing me to pursue this Ph.D. degree.

Secondly, I am grateful to all faculty and staff of the Computer Science and Engineering Department for their consistent support and willingness to answer all my questions. Specifically, I want to express my most sincere gratitude to my advisor Dr. Bill Buckles for guidance and patience. His kind and helpful supervision throughout the process were beyond my expectations, and I am deeply honored to have worked with him as a student. I believe I received immeasurable support from him, and I wish him and his lovely family good health and prosperity. To my committee members—Dr. Armin R. Mikler, Dr. Robert Akl, and Dr. Kamesh Namuduri—I would like to express my gratitude for being a part of this journey.

Most importantly, I want to thank all my family members. My parents , Md.Nazrul Islam and Mrs. Kohinoor Islam for bringing me in the world. I love you, ammu and abbu. Thank you ammu, for all of your sleepless nights that you have spent by my side and all of your sacrifices that you have made for your career to lift mine. Abbu, nobody knows better than me of your silent sacrifices, your hard work, and your singular vision for our family. I am fortunate to be your child. My degree is an endeavor to make you both proud. I would also like to express my special gratitude to my uncle, Dr. Mozammel Hoque for his stream of support and affection that has been flowing to me since my childhood.

Last but not least, thank you, my wife, Jennifar, for tolerating this bumpy life of a graduate student, and never faltering to be by my side.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
CHAPTER 1 INTRODUCTION	1
CHAPTER 2 LITERATURE REVIEW	5
2.1. Color Space based Skin Detection	5
2.1.1. Basic Color Spaces	6
2.1.2. Perceptual Color Spaces	8
2.1.3. Perceptually Uniform Color Spaces	9
2.1.4. Orthogonal Color Spaces	10
2.2. Skin Detection Methods	12
2.2.1. Pixel based Skin Detection	12
2.2.2. Boundary based Skin Detection Models	18
2.3. Clustering Techniques	20
2.3.1. Partitional Clustering	21
2.3.2. Hierarchical Clustering	24
2.3.3. Density based Clustering	24
2.3.4. Grid based Clustering	24
2.3.5. Spectral Clustering	25
2.4. Cluster Validity Assessment	26
2.4.1. Validity Indices	27
2.5. Discussion	29
CHAPTER 3 CLUSTERING BASED SKIN DETECTION	30
3.1. Clustering of Image	31

3.1.1.	Feature Selection	32
3.1.2.	Dynamically Assign Number of Clusters	34
3.1.2.1.	Label Binary Image	36
3.1.2.2.	Areas of Binary Image	38
3.1.2.3.	Area of Objects	40
3.1.2.4.	Number of Clusters	44
3.1.3.	Select Skin Cluster	46
3.2.	Region Growing	48
3.2.1.	Region Growing Algorithm	49
3.2.2.	Region Growing Results	51
3.3.	Summary	53
CHAPTER 4	EXPERIMENTAL EVALUATION	54
4.1.	Dataset Description	54
4.1.1.	Evaluation Metrics and Matrices	60
4.1.2.	Pixel based Skin Detection for Compaq Dataset	65
4.1.3.	Color based Skin Detection for Compaq Dataset	71
4.1.4.	Region based Skin Detection for Compaq Dataset	83
4.1.5.	Skin Detection for SFA Dataset	84
4.1.6.	Skin Detection for HGR Dataset	90
4.2.	Summary	94
CHAPTER 5	CONCLUSION	95
REFERENCES		98

LIST OF TABLES

	Page
Table 3.1. Missing skin pixels from clustering	50
Table 4.1. Compaq dataset details	55
Table 4.2. Example of Compaq dataset images	56
Table 4.3. SFA dataset details	57
Table 4.4. Examples of SFA dataset images	58
Table 4.5. HGR dataset details	59
Table 4.6. Example of HGR dataset images	59
Table 4.7. List of methods compared for Compaq dataset	66
Table 4.8. Comparison of accuracy and false positive rate of statistical methods	67
Table 4.9. Single color space based skin detection [57]	73
Table 4.10. Combined color space based skin detection [57]	76
Table 4.11. F-measure of different methods for various color spaces [42]	80
Table 4.12. Region based skin detection results for Compaq dataset [58]	83
Table 4.13. List of methods compared for SFA dataset	85
Table 4.14. SFA dataset experimental results [81]	85
Table 4.15. Experimental comparisons for SFA dataset [25]	88
Table 4.16. Experimental results for HGR dataset [32, 14, 5, 44, 73]	90
Table 4.17. HGR dataset results published by [62] for skin detection	92

LIST OF FIGURES

	Page
Figure 2.1. RGB, HSV, CIELAB, and YUV color spaces [52]	7
Figure 2.2. Taxonomy of skin detection method	13
Figure 2.3. Histogram of an image	15
Figure 2.4. K-means: possible initial centroid positions	22
Figure 2.5. K-means: first iteration	23
Figure 2.6. K-means: after two hundredth iterations	23
Figure 3.1. Overview of clustering based skin detection	32
Figure 3.2. CIELAB color space. (Uploaded in DOCPLAYER .Permission given for personal and non-commercial use)	33
Figure 3.3. Comparison of skin color distribution of RGB, HSV and CIELAB color space	34
Figure 3.4. Pixel value transformations of an image	35
Figure 3.5. An example of image labeling(only some parts are shown)	37
Figure 3.6. Objects or Blobs calculation from image labeling	39
Figure 3.7. Outer contour (in red) and inner contour (in green) of an image	40
Figure 3.8. Area of all connected components	41
Figure 3.9. Area covered by largest ten percent objects' in compaq dataset	42
Figure 3.10. Area covered by largest ten percent objects' in SFA dataset	42
Figure 3.11. Area covered by largest ten percent objects' in HGR dataset	42
Figure 3.12. Reconstruction of image from largest ten percent objects' area	43
Figure 3.13. Dynamically assign number of clusters	44
Figure 3.14. Histogram of hue value for non-skin and skin of compaq dataset	47
Figure 3.15. Histogram of hue value for non-skin and skin of SFA dataset	47
Figure 3.16. Histogram of hue value for non-skin and skin of IBTB dataset	47
Figure 3.17. Example to choose skin cluster	48
Figure 3.18. Calculation of skin seeding points	51

Figure 3.19.	Recover skin pixels by region growing	52
Figure 4.1.	Percentage of skin image types in compaq dataset	55
Figure 4.2.	Confusion matrix of skin detection	61
Figure 4.3.	Parameters of skin detection	62
Figure 4.4.	Confusion matrix example from an image	62
Figure 4.5.	Accuracy comparisons of parametric methods	67
Figure 4.6.	Accuracy comparisons of fusion based methods	69
Figure 4.7.	Accuracy comparisons of all statistical based methods	70
Figure 4.8.	False positive rate(FPR) comparisons of all statistical based methods	70
Figure 4.9.	Accuracy and FPR for all statistical methods	71
Figure 4.10.	Comparison of skin detection(proposed vs statistical methods)	72
Figure 4.11.	High and low recall and precision	73
Figure 4.12.	Recall, Precision, FPR, Accuracy and F-measure of single color space	74
Figure 4.13.	Overall performance of single color space	75
Figure 4.14.	Accuracy of combined color space	77
Figure 4.15.	FPR of combined color spaces	77
Figure 4.16.	Recall of combined color space	78
Figure 4.17.	Precision of combined color space	78
Figure 4.18.	Precision, Recall, and F-measure of combined color space	79
Figure 4.19.	Overall performance of combined color space	79
Figure 4.20.	F-measure for six color spaces of eight skin classifiers	82
Figure 4.21.	Region-based skin detection	83
Figure 4.22.	F-measure for six color spaces of eight skin classifiers	86
Figure 4.23.	Overall performance for skin detection using SFA dataset [81]	87
Figure 4.24.	Precision, Recall and F-measure of SFA dataset [25]	89
Figure 4.25.	Overall performance of SFA dataset [25]	89
Figure 4.26.	Precision, Recall and F-measure for HGR dataset [32, 14, 5, 44, 73]	91
Figure 4.27.	Overall performance of HGR dataset [32, 14, 5, 44, 73]	92

Figure 4.28.	Accuracy, FPR and FNR for HGR dataset [62]	93
Figure 4.29.	Overall performance of HGR dataset [62]	93

CHAPTER 1

INTRODUCTION

Skin detection is the most widely used primitive in human image processing. Various applications over images with human content employ skin detection as a primary step. Skin detection is the process of discriminating human skin pixels from non-skin pixels in an image or video [97]. Various human computer interaction (HCI) applications - face detection, face tracking [70, 48], gesture recognition [63], computational health informatics, web content filtering are heavily depend on efficient skin detection. Skin detection is also used as a cue for detecting people in general. The main challenge is to design skin detection methods robust to the large variations in appearance that can occur. However, there are various factors that make skin detection challenging. Among them, the notable one is variations in illumination. Variation in illumination is the most important factor that influence skin segmentation performance [77]. A variation of illumination (indoor, outdoor, shadows, presence of various light colors) produce changes in human skin tone. Shadows on the face from direct lighting may obscure some facial regions. It can be a reason for rendering a facial shape unevenly.

On the other hand, a bright facial region may be due to the natural diffusion of light causing extra brightness on the face. Additionally, various ethnic groups have many skin tones. Skin color tone varies from person to person within the same ethnic group and also between members of different ethnic groups. Caucasians, Africans, and Asians have different skin hues ranging from white, to dark to brown [79]. Presence or absence of shadows in an image or videos and complex background are another crucial parts of correct skin detection. Many real-world background objects have similar skin color. Skin color can be close to other image artifacts such as furniture, clothes, hair color etc. The diversity is virtually unlimited. This causes skin detector to produce false positive detection [38]. In addition image sources are an uncontrollable factor. ¹

¹Parts of this chapter have been previously published, from ABM Rezbaul Islam, Ali Alammari and Bill Buckles, "Skin detection in image and video founded in clustering and region growing", Proceedings Volume 10993, Mobile Multimedia/Image Processing, Security, and Applications 2019; 109930V (2019), Baltimore, Maryland, United States. Reproduce with permission from SPIE (<https://doi.org/10.1117/12.2518765>) [34].

Images available on the internet, newspapers may have been modified and filtered possibly multiple times. This makes color information problematic. Camera characteristics cause additional difficulty for skin detection. Because of these and other reasons, it is challenging to choose a specific trichromatic color space for any specific application. In addition, designing skin detection models to detect human skin under all possible constraints has not been achieved. Efficient handling of aforementioned challenges demands a model that is capable of differentiating skin and non-skin pixels. In this dissertation, a skin detection model is proposed which can overcome these challenges and perform better in real-world skin detection applications.

Researchers have been working to find a technique which will be able to detect skin in spite of challenges. The skin detection can be considered a binary classification problem. There are two principal types of skin detection pixel-based or region based. In pixel-based detection, a pixel is classified compared with its neighbors as either a skin pixel or not. Skin detection that is based on a specific trichromatic color space and specific color component value ranges is an example. On the other hand, region-based skin detection focuses on spatial arrangement of skin pixels with additional information of intensity and texture. However, some authors [84, 38, 56] contain the results of surveys on skin color modeling and skin segmentation based on color information in different color spaces. Phung [56] has discovered that skin detection accuracy does not depend on choice of color space. A few researchers have used texture [82, 19, 31] or shape [39] information in combination with color-based methods. Wavelets, contourlets [15, 35] are among the textural features that have been used in conjunction with color cues. Convolutional Neural Networks (CNNs) have received wide attentions more recently for image classification and object detection. But CNNs do not work well defined relationship between a pixel and its neighbors. Since skin pixels are located close to each other spatially, CNNs lack the desired performance outcomes.

Given all of the previous work, color cues have provided the dominant source of information. In spite of the emphasis on color-based analysis, a considerable number of applications will benefit from a system that can perform skin detection in the absence of color cues. More-

over, approaches mentioned above depend heavily on color information that, in the absence of global information, can lead to incorrect detection. Given the number of methods from which to choose, it is difficult to pick one which would project satisfactory accuracy under diverse conditions. So, it is worthwhile to seek a new technique for better performance. The purpose of this research is to find a skin detection method which does not rely on any predefined thresholds. A skin detection technique must be robust in spite of varying conditions. Moreover, the proposed technique must be tested and experimentally evaluated to establish its reliability. The primary contributions of the research are:

- Luminance of skin color is separated from skin tone by taking advantage of the broad spectrum of LAB color space. The proposed method considers only A and B components for skin detection. It ensures the pixels characteristics without being biased of luminance (lightness/darkness) value of an image.
- The main challenge for an unsupervised learning is to automatically define the number of clusters. Features in images vary greatly thus the number of clusters should be define dynamically. This challenge was solved in this research by doing just that.
- For region growing, seeding points are crucial. Hence, by over clustering, it can be ensured that chosen region seeding points are skin pixels. Moreover, using the standard deviation of skin pixels as the homogeneity predicate enables grouping maximum number skin pixels within each region.

In summary, the interaction between trichromatic color space (local information), clustering (global information) realized with a well-seeded region growing procedure leads to a system having superior accuracy and precision. Clustering based skin detection is a new approach compared with extant methods. The simplicity of computation and ease of feature selection of this approach can be adapted to any skin detection application. An efficient skin detection will elevate human computer interaction tasks to next level. Gesture recognition, face detection, filter objectionable content in web will be more straightforward and accurate. However, in unsupervised learning method clustering, it is problematic to automatically de-

fine number of clusters since the number of contained objects is unpredictable. The results of this research will open a new window of opportunities for researchers and practitioners interested in solving similar challenges that can be applicable to other types of problems.

The remaining of the dissertation is organized as follows:

Chapter 2 reviews various research works that is related to the research. Color space based skin detection which covers different color spaces used for skin detection is discussed.

Chapter 3 describes the overall clustering based skin detection method in details. All the steps related with the proposed method has been explained with relevant explanations.

Chapter 4 explains the experimental evaluation of proposed method. For evaluation some publicly available dataset was used and compared with many existing skin detection methods. Many evaluation parameters was considered for fair comparison.

Chapter 5 concludes the dissertation and direct to future works.

CHAPTER 2

LITERATURE REVIEW*

In this chapter, works related to this dissertation are discussed in order to more fully explain the reasoning behind aspects of the research explored in this dissertation. The first part is color space based skin detection. This is then followed by a discussion of different clustering techniques used to classify regions of the image. The last topic that is explored is research on the cluster validation index.

2.1. Color Space based Skin Detection

Color is not an innate property of an object; it is actually an emitted energy or reflection of light on objects. As the physical property of light, when light hits an objects, some of the wavelengths are absorbed and others are reflected. A human eye or camera sensor perceive these reflected wavelengths defines as the object's color [29].

In image processing, colors are transformed into numbers. The mathematical model that represents these images as numbers is known as color models or color space. The color space represents and visualizes colors as tuples of numbers, typically as three or four values of color components (or channels). Representing color in this way allows for a convenient method to identify colors [64, 74].

Color space is a 3D space with axes appropriately defined for all possible human perceptions. Among different color spaces, some of them are imaging device dependent and not close to human perception. These color spaces are often modified for digital applications. Although, color is not directly utilized in various skin detection methods, but it affects the performance of a skin detection algorithm. Albiol [3] has mathematically shown that skin detection performance is independent of the choice of color spaces. Device independent color spaces are popular among researchers for skin detection.

*Parts of this chapter have been previously published, from ABM Rezbaul Islam, Ali Alammari and Bill Buckles, "Skin detection in image and video founded in clustering and region growing", Proceedings Volume 10993, Mobile Multimedia/Image Processing, Security, and Applications 2019; 109930V (2019), Baltimore, Maryland, United States. Reproduce with permission from SPIE (<https://doi.org/10.1117/12.2518765>) [34].

In the field of skin detection, the most widely used color spaces are classified as follows:

- Basic Color Spaces (RGB, Normalized RGB)
- Perceptual Color Spaces (HSI, HSV, HSL, TSL)
- Perceptually Uniform Color Spaces (CIE-Lab, CIE-XYZ and CIE-Luv)
- Orthogonal Color Spaces (YCbCr, YIQ, YUV, YES)

2.1.1. Basic Color Spaces

RGB color space can be represented on the Cartesian coordinate, in each coordinate three components: Red (R), Green (G), and Blue (B) are described. It can be geometrically represented in a three-dimensional cube, as shown in 2.1(a). The RGB color space model can be implemented according to any skin detection method chosen. Combinations of the three primary colors define different colors in this model. In Figure 2.1(a) the corners of each axis represent the primary color component. RGB color space is an additive model, the origin of the cube starts from black, and intensity values are added for deriving other colors.

Brand, J and Mason [9], Jones and Regh [37], Caetano and Barone [11], Oliver [55] , Kim [43], Schwerdt and Crowley [68], Sebe [69], Storrington [77], Wang and Sung [86], Yang and Ahuja [91], Yang [90], use RGB color space for skin detection. However, high correlations between channels is a drawback of using this color space because it is unable to separate the chrominance and luminance portion of the image.

In the research conducted for this dissertation, RGB, even through its drawback of mixing chrominance and luminance, has advantages after becoming normalized. RGB is straightforward and, except for possible round-off errors loses no information [64]. The transformation of RGB to normalized RGB can be obtained by the process of normalization [64]. The transformation is done by the following (equation 2.1, 2.2, and 2.3)

$$(2.1) \quad R_n = \frac{R}{(R + G + B)}$$

$$(2.2) \quad G_n = \frac{G}{(R + G + B)}$$

$$(2.3) \quad B_n = \frac{B}{(R + G + B)}$$

Where R_n, G_n and B_n denotes the normalized R,G and B values respectively.

After running the image data through the above equations, all channel values are normalized. As the sum of all the channels is known, then the third component can be avoided. This technique reduces space dimensionality. The brightness of source RGB deeply depends on

FIGURE 2.1. RGB, HSV, CIELAB, and YUV color spaces [52]

the red and green channel, and normalization can separate it. Therefore, on a matte surface, the ambient light can be ignored, and normalized RGB remains constant for the changes in surface [72]. These advantages attract many researchers on color space based skin detection [10, 75]. Even normalized RGB color space suffers from uneven illuminations. Ultimately, the RGB color space (normalized or not) does not allow for RGB, the luminance channel to be separated from chrominance which results in a very tight pseudo-ellipse skin cluster. The real world conditions vary a lot, so tight pseudo ellipse skin cluster is not efficient.

2.1.2. Perceptual Color Spaces

HSV, HSI and HLS color spaces were an artist idea of using saturation, tint, and the tone. Perceptual color spaces such as HSI, HSV, HSB, HSL and TSL are very attractive color spaces in skin detection literature. In these color models, each pixel is presented by Hue (tint or color), Intensity (lightness) and Saturation associated with human perception of color [88]. These concepts introduce HSV color space. HSV color space can be characterized as a circular or hexagonal cone, as shown in Figure 2.1(b). In the HSV color space three components are Hue, Saturation and Value. Hue (H) defines the most dominant actual color of a specific area. Hue is represented as an angular measurement to a location around a color wheel (0^0 to 360^0) and the pure hues are along the outer edges of the wheel. The second component of HSV is Saturation and it represents colorfulness in relation to brightness. The value of saturation ranges is 0 to 1. The last component, value (also known as variance, intensity, or lightness) is the measurement of the lightness or darkness of a color area. In other words, how dark or light a specific color is. There is another color space similar to HSV called TSL (tint, saturation, and lightness), but HSV has gained more traction within the skin detection field of study.

HSV is an advantageous choice of a color space in skin detection because of its intuitive color components and it's ability to distinguish between luminance and chrominance. HSV color space is used by Garcia and Tziritas [26], McKenna [47], Saxe and Foulds [66], Sobottka and Pitas [73], Thu et al [33], Wang and Yuan [87], Zhu et al [96]. In [92], various characteristics of hue invariance to highlights at white light sources, matte surfaces to ambient

light, and surface orientation for light sources is discussed. Poynton [60] continues to discuss shortcomings such as, discontinuation of hue and brightness computation complexity with color vision properties. Moreover, logarithmic transformation based HS named as Fleck HS introduced by Fleck and used by Zarit [95].

2.1.3. Perceptually Uniform Color Spaces

CIE color space was introduced by CIE (International Commission on Illumination) system represent the color space with Y as luminance. CIE-XYZ was developed from a psychological experiment, and it is close to the human visual system [60]. Besides, various device dependent color spaces are used for TV transmission and digital photography. The striking difference between this color space and others is that CIE color spaces arrange all the colors by the perceptual difference of color; color, spaces-luminance (L) and chrominance are represented in XYZ coordinates. These coordinates are intended to correlate to the perceptual concept of human vision. So, therefore, this color space is efficient for skin detection. CIELAB and CIELUV are two examples of perceptually uniform color spaces. Perceptual uniformity defines a small distortion to a component value and can be equally observable across the range of values. RGB and other well-known color spaces are not perceptually uniform. Nonlinear transformation of perceptual color spaces corrects this shortcoming.

Perceptual color spaces need a complex transformation function for conversion. When an image is converted to CIELAB, it may appear precisely like the previous color space. The reason is CIELAB, and CIELUV retain all possible colors and because their translation of color is not related, there is no deterioration in image quality. It may be little tricky since in usual color spaces (e.g., RGB, HSV) contains logical colors. In CIELAB, one of the channels includes no color, and the other two channels have dual color combinations without any contrast. The L channel is for lightness, and there is no color value associated with it. It usually depicts the difference between the presence of darkness and light. In addition to this, A is the color balance between green and red, and B is a balance between blue and yellow. The conversion is done by the following (equation 2.4)

$$(2.4) \quad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490186 & 0.309879 & 0.199934 \\ 0.177015 & 0.812324 & 0.010660 \\ 0.010077 & 0.989922 & 0 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

where X,Y are two axes which represent wavelength of lights and Z holds luminance. R,G and B are three color channels of RGB color space. All of them have value range from 0 to 1. For aforementioned advantages Zarit et al [95], Yang and Ahuja [91], Schumeyer and Barner [67] used this color space in their research because of this color space's advantages.

2.1.4. Orthogonal Color Spaces

The orthogonal color space family includes well-known color spaces such as YCbCr, YCgCr, YDbDr, YPbPr, YIQ, and YUV. Due to the particular features of these color spaces: the separation of luminance-chrominance channels, relatively simple conversion, and relatively tight skin cluster, orthogonal color models have been frequently used in skin segmentation [26, 47, 36, 46].

YUV color space is used by PAL and SECAM TV broadcasting. Y component represents the luminance and U and V has all color information. For 8-bit image, the transformation of RGB to YUV color space is done by (equation 2.5) as described in [64]

$$(2.5) \quad \begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.500 \\ 0.500 & -0.418 & -0.081 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

where the range for Y is from 0 to 1 , while U and V range from -0.5 to 0.5, shown in 2.1(d).

YIQ color space is used by North America, Japan, Thailand, Korea and the National Television Systems Committee (NTSC). The most significant difference of YIQ from other color spaces is the separation of grayscale information from color data. The luminance is represented by Y (grayscale image), and IQ holds chrominance. The transformation of RGB

to YIQ color space is done by (equation 2.6) as described in [64].

$$(2.6) \quad \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}$$

Discussion on color space based skin detection: Color spaces affect the performance of skin detectors. Several authors experimented with color space choices in skin detection to see the results of their influence on skin segmentation algorithms performance. Vladimir [83] claimed that TSL is the best color space for skin detection when combined with Gaussian and Mixture of Gaussian models. In this research, he compared the performance of a variety of color spaces with a single Gaussian model which yielded YCbCr as having the best performance [41]. In another study, Montenegro et al [49] compared RGB, HSV, YCbCr, CIE-LAB and CIE-Luv implementing SFA data set. Matthews Correlation Coefficient (MCC) was used as the performance metric and Gaussian classifier as the method. This study concluded that, CIE-LAB outperforms all others and normalized RGB performs the worst. Furthermore, performance of RGB based techniques suffers in decimating skin-pixels while objects increases [3]. A comparison study is done using Gaussian and histogram approaches for a data set of 805 color images [35]. This works claimed that the choice of color space significantly changes the performance, and HSI color space accompanying with histogram model outperform others. In comparison of three different color spaces (HIS, RGB, CIE-LAB and YCbCr), in terms of classification error is obtained in HIS and YCbCr models [56]. A recent study [78] claims YPbPr outperforms other models. Besides Gaussian approaches, Gonzalez et al [16] compared the performance of ten common color spaces based on K-means clustering algorithm and conclude that HSV, YCgCr, and YDbDr are the best color spaces for skin detection. Nalepa et al [53] showed that statistically combination of

RGB and HSV outperforms other color spaces. In other hand, artificial neural network has been subjected for skin segmentation in variety of color spaces yielding to the conclusion that YIQ is the best choice as described by Al-Mahir [1]. The idea to adapt an optimal chrominance color space rather than a segmentation model has been suggest by Dahm [18]. Here a non-linear transformation between YUV and new TSL color space is used to boost the segmentation performance. In summary, the performance of color space based skin detection highly depends on various factors, including used methods. To make a fair comparison, all influencing factors should be considered. As observed, different authors take account of varying training, testing, and validation data. As a result, optimum color space for skin segmentation has changed to adapting optimal skin detection models. Besides, in neural networks, the performance of the classifier directly depends on the number of neurons, as well as the initial guess of the weights.

2.2. Skin Detection Methods

Skin detection processes can be organized into a collection of categories that are not mutually distinct. The central divide in skin detection methods are: Pixel based skin detection and Region based skin detection as seen in, Figure 2.2. Among pixel based methods there are three main types of methods-Statistical Based Method, Machine Learning Method and Threshold Based Method. Statistical strategies are solely based on records extracted from a histogram of training skin and non-skin pixels. Non-parametric or Parametric models are developed with the aim of acquiring the probability that a pixel belongs to skin or a non-skin class. Artificial neural networks (ANN) are very beneficial gear in both estimation of skin distribution or direct classification of the pixels. Adaptive techniques often reach better accuracy with the cost of computation. Support Vector Machine (SVM) based techniques also are used for classification of skin pixels in applications.

2.2.1. Pixel based Skin Detection

All pixel based skin detection work to classify skin pixels as skin or non-skin. These methods determine the properties of pixel based on color.

FIGURE 2.2. Taxonomy of skin detection method

In Statistical Models, skin detection is considered a probabilistic problem. There are many of techniques based on general distribution of skin and skin color that use specific color models in order to better detect skin. Based on an extensive training set, these strategies estimate the probability that a certain pixel is associated with skin. These methods use Histogram, Lookup Tables, Distance Based Detectors, Naive Bayes, and Gaussian Distribution.

Histogram based (Non-parametric) Models, in order to evaluate the data in skin detection it is important to plot it in a way that is beneficial to understanding in useful way. The Histogram Chart is useful tool to represent data graphically as shown in Figure 2.3. Each value that occurs in the data set has its own vertical bar. The shape of the histogram reveals the data's frequency of distribution. The frequencies displayed in the histogram are represented as probability distribution. The height of a bin in the histogram is proportional to the probability that the color is a skin color with (equation 2.7 and 2.8).

New pixels for which the corresponding likelihood value is greater than a predefined threshold θ are classified as skin pixels (equation 2.9), otherwise classified as background. Gomez [28] proposed a simplified probability theory using RGB color space.

$$(2.7) \quad P(rgb|skin) = \frac{Hist_{skin}[rgb]}{Total_{skin}[r, g, b]}$$

Where $P(rgb|skin)$ denotes the conditional probability of a pixel whether it is skin pixel.

$$(2.8) \quad P(rgb|\sim skin) = \frac{Hist_{\sim skin}[r, g, b]}{Total_{\sim skin}[r, g, b]}$$

Where $P(rgb|\sim skin)$ denotes the conditional probability of a pixel whether it is a non-skin. A new pixel is labeled as skin if it satisfies a given threshold θ .

$$(2.9) \quad \frac{P(rgb|skin)}{P(rgb|\sim skin)} \geq \theta$$

The main advantage of histograms are simple and easy to implement that give good performance. However, histogram charts cannot give any spatial relationship between pixels, which make it impossible to determine individual features of interest.

Lookup Tables(LUTs) is the fastest way for skin detection since it is not dependent on arithmetic operations. There are no specific definition of probability density characteristic in non-parametric techniques. Single histogram primarily based Look-UP-Table (LUT) model is a common approach in modeling skin color cluster. This technique, uses a set of training skin pixels from a distribution of skin pixels in the color space chosen for the skin detection program. The training technique is simple but populating the histogram in this way calls for an extensive skin data set.

Bayes Classifier Approach is based on Bayes theorem. Bayesian classifier considers two histograms skin, and non-skin pixels, from training data normalized to provide probability distribution. The conditional probability density function is given using Bayes rule

FIGURE 2.3. Histogram of an image

(equation 2.10):

$$(2.10) \quad P(skin|x) = \frac{P(x|skin)P(skin)}{P(x|skin)P(skin) + p(x|\sim skin)P(\sim skin)}$$

Where the conditional probability density function of a colored pixel, x is $P(skin|x)$. In addition, $P(skin)$ and $P(\sim skin)$ are the prior probabilities of skin and non- skin. $P(x|skin)$ and $p(x|\sim skin)$ are prior probabilities density of a given pixel x .

In Choi [17] and Zaidan [94], a Bayesian classifier based on the YCbCr color space is employed. Erdem et al [24] employed Bayesian based post filtering technique to reduce false positives of the Viola-Johns face detector. Zarit et al [95] performed a comparison study of overall performance of histogram method in Fleck HS, HSV, RGB, CIE Lab and YCbCr. Phung et al [56] examined the trade off among the wide variety of bins in line with channel and detection rate.

There are numerous gains and drawbacks associated with histogram based also known as non-parametric techniques. The benefit of this method include: training and implementation, rapid training and independency at the shape of cluster [84]. However, one major

detriment of this method is their dependence on the training set that requires the gathering of a massive range of skin pixels. Furthermore, large memories are required to implement histogram based models especially when fine resolutions are required.

Parametric Models include pattern classification by Gaussian distribution. This method is heuristic one and relies on skin color distribution that can be modeled based on Gaussian distribution in the color space. Most Gaussian distribution produces a symmetric, bell-shaped curve. The normal distribution depends on the mean μ and standard deviation σ . The mean determines the center and standard deviation determines height and width. The Probability Density Function (PDF) of a random variable x is given by $P(x)$ (equation 2.11) at Duda [22] :

$$(2.11) \quad P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Where x is a normal random variable.

Single Gaussian method (SGMs), Gaussian Mixture Models (GMMs), Cluster of Gaussian Models (CGMs), Elliptical Models (EMs), etc, are evolved to compensate for LUT's shortcomings of needing an excessive storage requirement. Similarly, they generalize thoroughly with a tremendously smaller amount of training set. In SGM, there should be a smooth Gaussian distribution around the mean vector. GMM is advanced version that has more complex distributions as a normalized weighted sum of Gaussian PDFs [91]. GMM compensates the inability of SGM in managing out of control conditions in a general skin segmentation problem. Similarly, SGM is not capable of approximating the actual distribution time due to the asymmetry of distribution in its peak [45]. The most appealing features of GMM models are their easy evaluation procedure and limited memory cost. The training technique is longer than the former methods. A comparative look at the overall performance of single and combination of Gaussian distributions in [12] confirmed that mixture method improve the performance only in a relevant operating area. However, skin color cluster have elliptical shape and use of Gaussian model leads to high false positive rate. Lee [45] proposed

a "elliptical boundary model" that outperform SGM. The model is defined in (equation 2.12)

$$(2.12) \quad \Phi(c) = (c - \phi)^T \Lambda^{-1} (c - \phi)$$

Non-parametric methods are primarily based on histogram of skin and non-skin pixels in a predefined training set. In comparison to parametric models, histogram methods are independent of the shape of cluster, trained efficiently. They can be utilized in several clock cycles per pixel needed for accessing memory. It was concluded that non-parametric models often outperform parametric ones with the cost of high storage requirement.

Machine Learning based Method(MLL), builds a skin detector that learns from a training data set and uses this knowledge for skin detection. This is called supervised learning. Skin classification can be classified into a skin or a non-skin pixel. One of the often used MLLs is, Artificial Neural Networks (ANNs) are mathematical models that stimulate the human nervous system. In skin detection, ANNs have been applied for specific functions and systems specifically in: illumination reimbursement, dynamic method, mixture with different strategies, and direct classification. A variety of ANNs such as multilayer perceptron (MLP), Self Organising Map(SOM),Pulse Coupled Neural Network(PCNN), etc., are commonly used. MLPs feeds forward synthetic artificial neural network. It consists of several layers of nodes in the acyclic directed graph where every layer is linked to the next one. Every neuron is a processing detail with a nonlinear activation feature besides for input ones. A common approach to educate MLPs is back propagation (BP), which is used at the side of optimization techniques along with gradient descent.

Support Vector Machine Model is a supervised learning method that applies to many pattern recognition tasks as well as human skin classification. During training, an annotated set of skin and non-skin pixels in given to the SVM for training. The SVM has a training algorithm that constructs a model that then attempts to assign pixels into the two training class,thus making it a non-probabilistic binary linear classifier. An SVM model

creates a representation of the pixels as points in space, mapped, so that skin and non-skin pixels are divided by clear and wide as a possible gap. New pixels are then mapped into the same space and predicted to fall in a category based on which side of the gap they fall on. Han et al [30] utilized an SVM based on active learning to detect skin pixels for gesture recognition, claiming that it is transferable to other applications.

2.2.2. Boundary based Skin Detection Models

The main difference in skin color is in its intensity (brightness) rather than the chrominance (color) [92, 16, 1]. Boundary based skin detection methods have a pixel-based processing scheme in which for any given pixel, methods are investigated to determine the class of that pixel. They are more prominent due to their efficient training, low-cost implementation and fast processing. However, several aspects are concerning researchers, particularly the degradation of classifiers, static nature, excessive dependence on training images. The effectiveness of rules, and inability to cope with maximum skin detection challenges. Kovac et al [44] proposed a method of explicitly defining a boundary model using RGB color space in two of daylight and flashlight situations which has been re utilized in [24, 50, 51, 59]. The boundary for skin detection in uniform daylight is mentioned (equation 2.13).

$$\begin{aligned}
 & R > 95 \ \& \ G > 40 \ \& \ B > 20 \ \& \\
 & (max\{R, G, B\} - min\{R, G, B\}) > 15 \ \& \\
 (2.13) \quad & |R - G| > 15 \ \& \ R > G \ \& \ R > B
 \end{aligned}$$

Orthogonal color areas are regularly utilized in case of explicitly described boundary models. Sagheer et al [65] exploited cases of regular lights and distinctive lighting fixtures. In order to detect faces, in [33, 26], rules are dynamically reconfigured based on pixels value. Zahir et al [93] proposed a boundary version using HSV color space for indoor and outdoor conditions (equation 2.14).

$$6 \leq H \leq 38$$

$$\begin{aligned}
9 &\leq S \leq 256 \\
100 &\leq V \leq 256
\end{aligned}
\tag{2.14}$$

Combinations of multiple color spaces has been effective in reinforcing the performance of explicitly defined boundary methods. CN Deshmukh [21] employed RGB, CbCr and HSV color method in which rules are set independently for triple color spaces and outcomes are clearly fused when taking final selection. The fusing of results has become a common procedure in reducing false positives to reduce false positives has been a common procedure; some examples are: RGB and YCbCr, YCbCr and YUV [82], HSV and YUV [85], HSV and YCbCr [40], RGB and YUV [89] and HSV and YCbCr [80]. The threshold values used by Vadakkepat [82] using YCbCr and YUV are as follows (equation 2.15).

$$\begin{aligned}
138 &< Cr < 38 \\
200 &< Cb + 0.6Cr < 215 \\
-30 &< U < 5 \\
-4.2 &< V < 28.8 \\
V &< 30 \\
U &< 0.45V - 37.65 \\
U &> -2.37V - 17.65 \\
V &< 30 \\
U &< 0.206V + 2.94 \\
U &> V^2 - 2.4V - 17.2
\end{aligned}
\tag{2.15}$$

Discussion : In order to carry out a fair and empirical evaluation of skin and skin segmentation strategies, it is essential to apply a standard and representative training and test set. From color space point of view, it seems that for that method in which the skin and skin cluster is greater compact, the specific policies are less complicated to design. But accuracy isnt always different in view that color spaces and consequent rules are convertible. The overall performance of Bayesian classifiers has been extraordinary, but also have an excessively high false detection rate. Additionally, the more memory is utilized for construction of the LUT; a better result is achieved. Bayesian methods when evaluated with SGM, GMM, MLP, and SOM are more efficient. Phung et al [56] compare the performance of multiple skin detection methods. In this study, the Bayesian classifier outperform other techniques with a sufficient difference. Due to the field evaluation nature of multi-spectral methods, it is not currently possible to compare the performance of these techniques either among themselves or with other systems. However, the high accuracy of such systems in most of normal situations is not questionable.

It is tough to derive a completely precise conclusion on the best method due to the wide variability of data from these sources. The Bayesian classifier generally outperforms, but has a wide variety of bins and a huge training set , but ultimately great classifiers in terms of accuracy. In terms of speed, computations and implementation cost, there is a trade off among strategies. With developing new strategies and techniques in recent years, former techniques are set apart, whilst precision gets first priority. These older methods are much slower than most of traditional methods which makes them incompatible for real-time applications.

2.3. Clustering Techniques

In this study, a clustering method is used; for that reason, clustering methods are discussed here. Clustering is a division of data into groups of similar objects. In each group, named a cluster, there are objects that are compared and grouped based on identical and dissimilar properties. If data is represented in the cluster, it will lose some fine details but achieve simplification. Most data clustering problems are considered NP-hard. Clustering

methods can be categorized into different paradigms: Partitional clustering, Hierarchical clustering, Density-based clustering, and Spectral clustering.

2.3.1. Partitional Clustering

As the name suggested, in partitional clustering, data is divided into non-overlapping subsets such that each data instance is assigned to exactly one subset. For example, K-means [76] and K-medoids are the most common method of its kind. K-means clustering applies an iterative approach. At first, it chooses the means of cluster commonly known as centroids. It then assigns data points to the nearest centroids. This approach is efficient in terms of computational speed and simplicity in use [71]. However, the vulnerability of the random seeding technique makes this method undesirable. If these initial seeding points are not chosen carefully, the result will not be satisfactory. For this reason, an updated method name K-means++ [4] was proposed to improve this unsatisfactory result. K-mediod is an improvement of K-means ability to handle discrete data. This improved method marks the most centralized data points as the representative of the corresponding cluster.

K-means Clustering, in this research, K-means clustering, is used for clustering. K-means is a clustering method which divides data points into a predefined number of disjoint clusters. The "K" refers to the number of clusters that are defined prior to clustering. To decide which data points belong to a specific cluster, various distance parameters are considered. For this study, a city block (equation 2.16) distance metrics is used:

$$(2.16) \quad CityBlockDistance = \sum_{j=0}^k |a_j - b_j|$$

where "k" denotes the number of dimensions and a_j, b_j are respective data points.

K-means algorithm partitions given data into a predefined number of clusters. The algorithm begins with a random set of K center points σ . For each iteration, data points are assigned to nearest center points (equation 2.17). To decide the nearness, various distance measurements are used. Usually, only one assignment to the center is possible. If the distance

FIGURE 2.4. K-means: possible initial centroid positions

between the center point is equal, then any one of the measurements are chosen at random. Initially, based on the number of clusters, random centroid values are assigned. It is shown in black at Figure 2.4.

$$(2.17) \quad S_i^{(t)} = \{x_p : \|x_p - \sigma_i^{(t)}\|^2 \leq \|x_p - \mu_j^{(t)}\|^2 \forall j, 1 \leq j \leq k\}$$

Afterwards, the center-points are re positioned according to the mean between the observation and respective center points (equation 2.18).

$$(2.18) \quad \sigma_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{x_j \in S_i^{(t)}} x_j$$

The update of the cluster will continue until all the observations remain unchanged to center points and when center points also remain unchanged. As stated before, K-means algorithm attempts to optimize the objective function (equation 2.19). However, the number of observations and centroids are finite so for each iteration it produces better clustering results.

Figure 2.5 shows the initial cluster centroid positions after the first iteration. Figure 2.6 depicts the centroid positions after two hundredth iterations.

FIGURE 2.5. K-means: first iteration

FIGURE 2.6. K-means: after two hundredth iterations

$$(2.19) \quad J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \sigma_k\|^2$$

$$(2.20) \quad \text{with } r_{nk} = \begin{cases} 1 & x_n \in S_k \\ 0 & \text{otherwise} \end{cases}$$

The main disadvantage of K-means is the performance highly depends on initial cluster centroids. A wrong choice of centroids can possibly split data that belong together can cause some centroids to be more inclined to an outlier.

2.3.2. Hierarchical Clustering

In this approach, clusters are formed by either a bottom-up approach or a top-down approach. This dataset is represented by a dendrogram. At the initial level, clusters are formed along the hierarchy. Single-linkage clustering is a classic bottom-up approach in which data points are gradually agglomerated together to form clusters. For each step, all the pairwise distances are calculated to find a minimum. The points with minimum distance are linked together. Steps are repeated until all points are linked. This results in a tree structure, and the depth of the tree can be used from cluster. Moreover, hierarchical clustering is suitable for the data set with different shapes and attributes of arbitrary type. However the time complexity is relatively high, and several clusters are needed to be preset. Some mentionable algorithms for this category are BIRCH, CURE, and ROCK.

2.3.3. Density based Clustering

Density based clustering is one of the well-known clustering techniques. This technique uses connectivity and density functions for clustering. In density-based clustering, data points are separated based on density, connectivity, and boundary. A cluster is formed at the base of the connected dense component and then is expanded to the direction of density. This property gives density-based clustering the ability to discover arbitrary shapes and remove the outliers. DBSCAN, OPTICS, DBCLASD, and DENCLUE are algorithms that are commonly used clustering techniques. The DBSCAN algorithm uses density-based notions to define clusters. It utilizes two connectivity functions, density-reachable and density-connected to define each data point as either a core point or a border point. DBSCAN then visits points arbitrarily until all points have been visited. However, if the point is a core point, it attempts to expand and form a cluster around itself. Based on the experimental results, it tends to derive cluster robustness toward discovering arbitrarily shaped clusters.

2.3.4. Grid based Clustering

In grid-based clustering, the data points are divided into multiple portions (grids) at different granularity levels to be clustered individually. This method's processing time is

significantly faster because it only concentrates on statistical properties of certain grid portions. After calculating all the grids, the results are consolidated and analyzed. This approach makes it independent in performing clustering on the grid rather than the whole data set. This technique is highly grid dependent, and the size of the grid plays a vital role in performances. For highly irregular data distribution, only some grids may not be formed adequately to achieve correct result. A few well-known grid based clustering methods are: Wave Cluster, STING, and CLIQUE. In particular, CLIQUE is written so that it can automatically find subspaces with high-density clusters and it does not require any assumption of data distribution. Its experimental results prove that this cluster can scale dimensions efficiently and, thus, it is exceptionally efficient in clustering high dimensional data.

2.3.5. Spectral Clustering

Various existing clustering methods uses an iterative approach to achieve a local minima and finds functional clusters using different initial cluster starting points. In contrast, spectral clustering uses maximum eigenvectors that are provided from the distance matrix as a form of a matrix. Spectral clustering usually reduces the dimension of K-means clustering. It considers the range of the similarity matrix of the input data. These clustering methods treat the cluster as a graph partition. For that reason, it does not make any strong assumption for the statistical properties of a cluster, therefore, in terms of time complexity, it is very fast. Its drawbacks, however, are that it can be sensitive to the choice of parameters and computationally expensive for large dataset.

Discussion: Partitional algorithms apply mainly to numerical datasets. However, a variation of K-means named as K-mode has been developed for categorical data. Moreover, they are based on a certain assumption to partition a dataset. Thus, they need to specify the number of clusters prior. The results of clustering are devoted to similar points that belong to the same cluster. Those points can be centroid or mediods, which depend on the clustering algorithm used. However, regardless of algorithm choice, the object remains the same- Minimization of distance within the same cluster and maximization of distance among separate clusters. Also, a centroid of a K-means represents the minimum distance

of data points that belongs to the same cluster. Hierarchical clustering methods create a hierarchical decomposition of the database represented as a dendrogram. This algorithm is a more effective method for detecting outlier and handling noisy data. BIRCH uses a hierarchical data structure called CF-tree for multi-phase clustering. In BIRCH, the first scan determines a good clustering and more scan improves the cluster quality. But it is only applicable for numerical data. Besides, it is an order sensitive, which means, for the same data points, it can demonstrate multiple clusters but in a different order. However, it has performance limitations for un-uniform data size and shape. The reason is, it heavily depends on the centroid value when data points are redistributed at the final stage. So, cluster quality is parameter depended. Density-based clustering algorithms suitably handle arbitrarily shaped collections of points as well as clusters of different sizes. Moreover, they can efficiently separate noise. Grid-based algorithms perform well for large databases. It can also find various arbitrary shapes of the cluster with the detection of outliers. A well-known grid-based algorithm is STING.

2.4. Cluster Validity Assessment

After clustering, these results must be evaluated, and this process is called cluster validity. This process requires the partition, to be evaluated to if it represents an accurate detection of skin. In general terms, there are three approaches to investigate cluster validity. The first is based on external criteria, which implies that the results of a clustering algorithm are evaluated based on a pre-specified structure. The second approach is based on internal criteria, where the cluster is assessed based on various proximity matrices. The last approach uses a relative measurement and involves. Comparisons of clusters to one another. A result of the same cluster, while looking at it with different parameters.

There are two criteria proposed for clustering evaluation and selection of an optimal clustering scheme [6]. The first, compactness, depicts how close two clusters are with each other. A standard measure of compactness is variance and this measurement should be minimized. The other one is separation and it measures the difference between the two clusters. Some common approaches are measuring the distance between two different clusters. Single

linkage, it measures the distance between the closest members of the clusters (equation 2.21).

$$(2.21) \quad d_{single}(G, H) = \min_{i \in G, j \in H} d_{ij}$$

Where G denotes the set $G = \{i_1, i_2, \dots, i_t\}$ of indices points of the group and H is the dissimilarity score. However, single linkage suffers from chaining problem. In other words, if two groups are need to be merged it needs one pair of close points avoiding others. As a result cluster can be very spread and not well compact. Complete linkage, it measures the distance between the most distant members (equation 2.22).

$$(2.22) \quad d_{complete}(G, H) = \max_{i \in G, j \in H} d_{ij}$$

Complete linkage avoids the aforementioned chaining problem but it suffers from crowding. In addition, its score H is based on worst case dissimilarity among pairs. So, the result can be compact not separate distinguished enough.

Comparison of centroids measures the distance between the centers of the clusters. These two approaches are based on statistical tests and their major drawback is their high computational cost. On the other hand, average linkage aims at finding the best clustering scheme that a clustering algorithm can be defined with certain assumptions and parameters.

2.4.1. Validity Indices

Quantitative evaluation of clustering results are known as cluster validity methods. There are three criteria-external, internal, and relative criteria. In an external criteria approach, data points examined to conclude whether or not they are randomly structured. This analysis is based on the Null Hypothesis expressed as a statement of the random structure of a dataset. This hypothesis is based on statistical tests, which lead to a computationally complex procedure. For internal criteria, the aim is to evaluate the clustering result of an algorithm using only quantities and features inherent to the dataset. Statistical testing is

the basis of the above-mentioned validation methods. Thus, the main setback is, since it needs consideration for internal or external criteria, it needs a large computation. For this, a different validation approach-relative criteria are proposed. It is based on relevant criteria excluding the requirement of any statistical tests. Ultimately it chooses the best clustering approach for the prior defied scheme based on the criterion specified.

However, the clustering method that is adopted in this study K-means clustering. It is an example of crisp clustering since the clustering object may have a membership of either skin or non-skin category. To evaluate the crisp clustering the well known Dunn and Dunn-like indices is used. A cluster validity index for crisp clustering was proposed in Dunn [23], attempts to identify compact and well separated clusters and gives the number of clusters. The index is defined as (equations 2.23, 2.24, 2.25) for a specific number of clusters.

$$(2.23) \quad D_{nc} = \min_{i=1...nc} \left\{ \min_{j=i+1...nc} \left(\frac{d(C_i, C_j)}{\max_{k=1...nc} \text{diam}(c_k)} \right) \right\}$$

Where, $d(C_i, C_j)$ indicates the dissimilarity function between two clusters C_i and C_j defined as:

$$(2.24) \quad D_{nc} = \min_{x \in c_i, y \in c_j} d(x, y)$$

$\text{diam}(c)$ denotes the diameter of a cluster, which may be considered as a measure of dispersion of the clusters. The diameter of a cluster C can be dened as follows

$$(2.25) \quad \text{diam}(C) = \max_{x, y \in C} d(x, y)$$

2.5. Discussion

- There does not exist any sole or mixed color space that can be used for better skin detection under various real world challenges.
- Statistical based methods require a large training data set and large memory.
- Researchers have taken different data sets and varying number of training and testing data which makes it difficult to compare existing approaches under a common standard.
- Clustering based methods with a selection of correct features can significantly improve the skin detection performance.

CHAPTER 3

CLUSTERING BASED SKIN DETECTION

The image processing field is a study that has many facets that are specific to it, such as segmentation, color coding, motion analysis, and many others. It also has aspects for which the scientific principle are imported from other areas of study. Of course, image processing can be a valuable test bed for developing new principle and analytic methods in which are applicable to various machine learning applications, data reduction, classification, and pattern recognition. In this research, a new technique in skin detection is explored, researched, and proven as efficient.

Human skin detection has various applications in Human-Computer Interaction (HCI), gesture analysis, facial expression detection, tracking of the face, and so on. Efficient recognition of skin regions within an image can define a specific region which then alleviates the issue of considering unnecessary parts in an analysis. The main goal of skin region detection is to detect skin pixels in an image and generate skin areas within the interested regions.

Moreover, the generated region is used for the desired applications. If this region is not correctly detected, then the applications based on these regions are incorrect. In addition to the issue of accurately detecting important areas of the photo, a skin based region detection can miss skin pixels or considers non-skin pixels as skin and therefore the reliability of various applications will decrease. Skin detection is a crucial initial step for a successful HCI based system. Numerous skin detection methods have been proposed and discussed briefly in this literature review, but their performances are often bounded by skin type and other surrounding factors. Till date, no skin detection method can perform efficiently for all types of skin color under varying real-world conditions.

¹This chapter is presented in its entirety from ABM Rezbaul Islam, Ali Alammari and Bill Buckles, "Skin detection in image and video founded in clustering and region growing", Proceedings Volume 10993, Mobile Multimedia/Image Processing, Security, and Applications 2019; 109930V (2019), Baltimore, Maryland, United States. Reproduce with permission from SPIE (<https://doi.org/10.1117/12.2518765>) [34].

Among various approaches [2] is the closest to this study’s proposed method. Two significant steps are involved in training an MLP ANN for higher detection accuracy and enhance skin detection using K-means and MLP ANN. Also, a set of input features are selected by differential evolution (DE). DE algorithm is used to select optimal input variables based on color and texture information. The six different texture descriptors for skin detection are: uniformity, stand deviation, skewness, kurtosis, smoothness, and entropy. After DE chooses input variables, the K-means clustering method is used to enhance the output of the optimized MLP ANN. This dissertation uses clustering for skin detection and differential evolution method for various blocks of the data set. Then the selected skin patches are used for MLP ANN training. Selecting skin patches are a difficult procedure. Current proposed clustering methods can separate most of the skin pixel from the image without DE. In this research, a cluster-based approach is proposed which will be used for skin detection and based on the skin detection, region for skin related application will be defined. The overall method is described in Figure 3.1.

3.1. Clustering of Image

Cluster analysis is a process of partitioning data into that have similar characteristics. Various components can be taken as a feature with metrics of counting similarity in the process of clustering. The clustering method is an efficient technique, especially for skin detection, because human skin is homogeneous. Regardless of ethnic group, all skin pixels and skin region show a substantial homogeneity. Furthermore, even when the background region has very similar skin color, there always exists a boundary between the true skin region and the background and this uniformity is not affected by lighting and other features of images. When an image is converted to a binary image, it is given an intensity value of 0 or 1. The background is always converted to black pixel and skin image in white. Clustering is an unsupervised approach, therefore, the data does not have any target attributes, but some intrinsic structures of data can be obtained. In this study, K-means clustering is adopted. For feature selection, CIELAB color space was taken with city block as a similarity metrics

FIGURE 3.1. Overview of clustering based skin detection

3.1.1. Feature Selection

To cluster images, feature selection is a crucial one. For this dissertation, CIELAB color space is used. In CIELAB color space, the L represents luminance (lightness) and it has a range from $[0-100]$. The A and B is for its two color channels. The A channel represents color in the range $[-128 \text{ to } +127]$ (green to red). The B channel has the same range $[-128 \text{ to } +127]$ (blue to yellow). A visualization of the LAB color space is in Figure 3.2. A crucial aspect of the color space is that the L component is the only channel that contains luminance. This allows for a pixel to be darker and whiter without changing the A and B channels. This is a unique characteristic compared to RGB and other color spaces. For skin detection, CIELAB is an efficient color space because the skin density in CIELAB space is

narrow for every ethnic group. For this narrow spread of skin density, more accurate skin detection is possible.

FIGURE 3.2. CIELAB color space. (Uploaded in DOCPLAYER .Permission given for personal and non-commercial use)

The Figure 3.3 displays: an image that only has skin pixels and three columns of the skin color distribution in RGB, HSV, and CIELAB. skin color distribution for three color spaces. The image has only human skin. All R,G and B channels of RGB color space has a distribution for skin color in the range of $[0-255]$, For HSV the distribution for saturation and value is spread all over the spectrum. But in CIELAB color space, a component has a range between $[0-30]$ and b has a range from $[0-30]$ approximately. Similar results was shown for various skin color of ethnic group. CIELAB's small range of skin color distribution makes for a better choice of color space allowing for more accurate skin detection.

For clustering purposes, only A and B elements were used. This is in order to allow for clustering without illumination because illumination often interferes with skin color properties. Ultimately, the A and B components contain all color information necessary for skin detection. This separation of luminance from the color channels makes CIELAB color space a good choice for skin detection compared to the other color space.

FIGURE 3.3. Comparison of skin color distribution of RGB, HSV and CIELAB color space

3.1.2. Dynamically Assign Number of Clusters

After selecting the right color space, the next challenging task is to define a number of clusters. As clustering is an unsupervised learning method, it needs to define the number of clusters before clustering. But the problem is image data varies in huge deviations; it is not realistic to have a fixed number of clusters for efficient skin detection. Each image has different color characteristics, and it needs to define a number of clusters dynamically.

An image is constructed with pixels. Thus the relationship between pixels reveals important information. This dissertation's approach takes an RGB image and converts it to grayscale. This shrinks the number of channels, but keeps the necessary reserve information. RGB image or color map conversion to grayscale eliminates the hue and saturation information while keeps luminance. This conversion is done by the following equation 3.1 where a weighted sum of R, G and B components are used [84].

$$(3.1) \quad GrayScaleImage = 0.2989R + 0.5870G + 0.1140B$$

Ultimately, the goal is to find the significant objects in the image in order to define the number of clusters. To cluster an image, it is crucial that the number of clusters should not be too high so that the image is clustered into too many unnecessary clusters. There should

be a right balance between the image contents and the number of clusters. Also, the image can be comprised of various objects and their shape; size can vary in an unknown number of possibilities. So, a shape based or face based object detection will not deliver a correct number of clusters since human presence in an image can be large or miniscule based on the orientation of image. In terms of pixel value, the area of the objects in the i will give an idea about the number of a significant number of objects in an image. The word significant means here the objects that have the biggest area, and it holds the most number of pixels in an image. To detect objects from the image, binary images are well researched and widely used since the pixel intensity value of a binary image is only 0 and 1. The assumption is, the intensity of pixels can be used to differentiate between background (black or white) and foreground. The intensity of an image depicts the importance of the pixel. Converting grayscale to the binary image is done by the threshold method. The output binary image has values of 1(white) for all pixels in the input image with luminance greater than threshold and 0 (black) for all other pixels. Hence, in the binary image, there are only two possible values. These values represent the foreground and background in Figure 3.4(c).

FIGURE 3.4. Pixel value transformations of an image

To find a number of objects present in the image, the first step is to group pixels that belong to the same objects. Hence, an object can be separate from touching foreground pixels, in another word a connected binary region. To get the binary region, each pixel needs to define to which region it belongs.

3.1.2.1. Label Binary Image

After creating the binary image, it needs to be given a label for each pixel. The labeling is done for connected components in the binary image. Labeling algorithm returns a matrix L with the same size as the binary image. The elements of L are integer values greater than or equal to 0. Pixels labeled as 0 are the background. The pixels labeled as 1 make up one object, the pixels labeled 2 make up a second object, and so on. The connected component describes a set of connected pixels that share the same properties.

Definition 01: A pixel $a \in S$ is said to be connected to $b \in S$ if there is a path from a to b consisting entirely of pixels of S . The term "path" is meant to be used here as ordered sequence of pixels in which any two adjacent pixels in the sequence are neighbors.

Moreover, a component labeling algorithm finds all connected components in an image. Then it assigns a unique label to all the pixels that belongs to the same component. During this process, neighboring pixels are pieced together in a stepwise manner to build regions in which all pixels within that region are assigned a unique number (label) for identification. An example is given in equation 3.2.

$$(3.2) \quad BinaryImage(u, v) = \begin{cases} 0, & \text{a background pixel} \\ 1, & \text{a foreground pixel} \\ 2, 3, \dots, & \text{a region label} \end{cases}$$

An example of component labeling is given in Figure 3.5. Component labeling as described in Algorithm 1 detects all the connected component. If there is an object in the image, then

all pixels representing that objects are connected. Noisy pixels are often generated because of discontinuity of background, infrared light reflectively, and fast movements. Besides background pixels are also connected and belongs to the same object. Binary large object (blob) refers to a group of connected pixels and the region can be bright or dark compared to the neighborhood. A number of blobs means a significant number of objects are present in the image. Separation of background and foreground are done with the flood fill operation.

Flood fill is especially useful when an area is defined with multiple color boundaries or irregular shapes. This algorithm searches for an unmarked foreground pixel and then visits and marks all the neighboring pixels in the region. A region is selected based on pixels with connectivity. In this research, 8-connected pixels were used, which means an unbroken path of 8-adjacent pixels connection. Flood fill uses local variables that are allocated by the recursive calls. Within each region, a tree structure rooted in the starting points with neighboring pixel relations.

FIGURE 3.5. An example of image labeling(only some parts are shown)

<p>Input : Binary Image</p> <p>Output: Label matrix of connected objects</p> <pre> 1 Binary image (u, v) = 0: background 2 Binary image (u, v) =1: foreground 3 The Binary image is labeled and returned 4 L←2 Value of the next label to be assigned 5 for <i>All image coordinates (u, v)</i> do 6 if <i>Binary image I(u, v) = 1</i> then 7 FloodFill (Binary image, u, v, L) 8 L←L+1 9 else 10 Go to the next Index 11 end 12 end 13 Return the labeled Binary image (I) 14 Exit </pre>
--

Algorithm 1: Component Labeling Algorithm

3.1.2.2. Areas of Binary Image

To get all areas of the binary image, each connected components (object) in the binary image which must be a logical array with any dimensions. The labeled region in the label matrix L. Positive integer elements of L represent different regions. For example, the set of elements of L having value 1 correspond to region 1; the set of elements of L having value 2 correspond to region 2; and so on. In this problem domain, the only area of the labeled regions is required. Moreover, the perimeter should be used on a label matrix or binary image with contiguous regions. Contiguous regions are also known as connected components, objects, or blobs. A label matrix containing contiguous and discontinuous regions might look like Figure 3.6(a),(b).

FIGURE 3.6. Objects or Blobs calculation from image labeling

In Figure 3.6 shows an example of calculating connected components. as shown in Figure 3.6(a), elements of L equal to 6 belong to the one object or connected component, elements of L which has the value of 8 belong to another connected component. In contrast, discontinuous regions in Figure 3.6(b) are regions that may contain multiple connected components. Elements of L equal to 6 belong to the first region, which is discontinuous and contains two connected components. Elements of L equal to 8 belong to the second region, which is a single connected component.

Once the region of a binary image has been found, the next step is to get the contour of the region. The pixels along the edge of the binary image are detected with morphological operations. There are two types of contour possible: outer contour and inner contour. Outer contour can only be formed with connected components. Due to holes, there are many inner arbitrarily contours are expected. These small holes indicate smaller regions among the bigger ones. To get only bigger blobs, these small holes are ignored since they do not contain any significant features.

It is obvious that skin pixels of mouth hole, teeth, eyebrows, nose hole, eyebrows are not considered for skin detection. It is always excluded from the computation. While contour of a human face or body is used for skin detection, it is essential to get rid of those small, unwanted regions since they do not contain measurable skin pixels. To obtain the exterior and interior boundary of an object, it is important to note that the relationship between them is known as parent and child. The outer most boundary is known as parents and the objects completely enclosed by parents are knows as children. Figure 3.7(b) describes the inner and outer contour of an image. The green color labeled "inner contour" or "hole" is insignificant for skin detection.

(a) Original Image

b) Outer Contour (red) and Inner Contour (green)

FIGURE 3.7. Outer contour (in red) and inner contour (in green) of an image

3.1.2.3. Area of Objects

After getting rid of insignificant objects, the area and centroid of the image blobs are calculated. The area is calculated with the actual number of pixels in the regions and centroid declares the mass of the region. The area of a binary region R can be defined by simply counting a number of pixels that consists in the region. Once all areas of the regions are calculated, this area defines the number of possible objects in an image. The number of clusters can be calculated more precisely for better clustering performance. The number of clusters should be based on the number of significant objects present in the image because significant objects will have the most number of pixels in the image. After reducing the number of objects by removing insignificant ones, the area and centroid of the image objects are calculated. In any image, many small connected components are visible. These small areas are so small that they rarely have any feature information. These objects are mostly hairs, ornaments etc. Figure 3.8(a), has a total of 18 connected components. In Figure 3.8 (b) the pixel counts for each connected component is given. It is quite obvious that, most of the area is really small (e.g., 1, 5, 10, 18 pixels only) as shown in Figure 3.8(c). It shows that only a small number of regions will contain most of the image pixels and rest of them

can be ignored since their numbers are insignificant.

FIGURE 3.8. Area of all connected components

In order to fully understand the scope of the problem in separating objects from one another in the image, some other observations need to be considered. For most of the image, only one or two objects contain most of the image pixels rather than a uniform distribution. Also, it can be more refined with an assumption that, among all objects counted, only the largest ten percent objects are considered since the area of the largest ten objects contains most of the image pixels. Due to the uncertainty of object appearance in an image, an experimental method was performed to establish the claim.

As described in Figure 3.9, 94% images of total 3775 of Compaq dataset, got more than 90% total pixels covered by the largest ten percent objects area. Only 3% of images have 60% total image pixels covered. Also to SFA data set in Figure 3.10, 99% images among 1118 were covered by more than 90% of total pixels. Lastly, for HGR data set in Figure 3.11, 94% images of total 898 were covered more than 90% total pixels. However, to prove the claim, it is also required to see which portions of the image pixels are covered. Skin pixels are homogeneous, and skin regions are mostly connected and consider as a single object. An object in an image contains a significant amount of pixels. Moreover, for images containing

FIGURE 3.9. Area covered by largest ten percent objects' in compaq dataset

FIGURE 3.10. Area covered by largest ten percent objects' in SFA dataset

FIGURE 3.11. Area covered by largest ten percent objects' in HGR dataset

multiple persons in an image usually presented with connected pixels and considered as one object. If it is possible to see the effect of this claim with the original image reconstruction from the largest ten percent objects, that will prove it undoubtedly.

FIGURE 3.12. Reconstruction of image from largest ten percent objects' area

In order to illustrate more fully the idea of 90% of the image is composed of 10% percent of the major objects' area, see Figure 3.12. It shows an example of image reconstruction from the largest ten percent objects area. It is important to note that so many of the images have almost all their pixels and return practically same image in reconstruction. In this Figure 3.12(a) only images where pixels are missing are shown. On Figure 3.12(b) the pixels that were missing are shown in light green color. Reconstruction of image is an experimental proof that, almost all the image pixels are can be reconstructed from largest ten percent

objects' area.

3.1.2.4. Number of Clusters

The number of objects that have singular properties. The next step in this process is to create the number of clusters and the problem lies in that the number of clusters must be defined. In this study, to dynamically define the number, the calculation of significant objects is taken by using the mean area of the largest ten percent of objects in the image. Then the mean area of those largest ten percent objects is calculated and the objects that have a lesser area than the mean value are dropped and the number of objects left is chosen as a number of clusters. As shown in Figure 3.13(a), there are two objects that have greater than average area, so number of clusters is two. For Figure 3.13(b), there are three objects with a greater area than average, so the number of clusters is three. However, if there is only one object, then for minimum clustering, a number of clusters are assigned as two.

(a) Number of clusters is 2

(b) Number of clusters is 3

FIGURE 3.13. Dynamically assign number of clusters

```

Input : Input Image(I)
Output: Number of clusters
1 Convert Input image to Grayscale(IG)
2  $IG(i,j) = [0-255]$  where  $i$ = number of rows and  $j$ = number of columns
3 Convert Grayscale to Binary (IB)
4  $IB(i,j) = [0 \text{ or } 1]$  where  $i$ = number of rows and  $j$ = number of columns
5 Label Binary Image
6 foreach pixel of IB do
7 |   Get a label matrix L with 8 connected object
8 |   where  $L = 0, 1, 2, 3$  number of objects
9 end
10 Get all the regions of Binary Image
11   Region  $\leftarrow \Phi$ 
12   Area  $\leftarrow 0$ 
13 foreach label region in L do
14 |   foreach same elements of L do
15 | |   Region = Region+1
16 | |   Area = Number of pixels in Region
17 |   end
18 end
19 Consider largest ten percent regions of Binary Image
20 Area1  $\leftarrow$  SortArea in Descending Order
21 Final Area  $\leftarrow$  Area1[1, 2, 10]
22 Get the mean area of selected regions
23 Mean Area  $\leftarrow$  Average(FinalArea)
24 Remove all regions smaller than average area
25 if Averagearea < FinalArea then
26 |   Remove the Region
27 else
28 |   Add the Region
29 end
30 Get the boundaries of selected regions
31   Trace exterior boundaries of selected region
32 Count Number of objects found in selected Region
33 Number of clusters = Number of object found
34 Exit

```

Algorithm 2: Select Number of Clusters

The whole process described previously is detailed in Algorithm 2. This process dynamically assigns a number of clusters based on object properties of each image and remove the bottleneck of prior assigning a number of cluster in the unsupervised learning method. The effectiveness of this process can also be applied to other applications.

3.1.3. Select Skin Cluster

After computing the a number of clusters needed in the skin detection process, the input images are converted to CIELAB color space. L component of images are then ignored because it contains only the lighting information which is unnecessary (and often detrimental) for skin detection. The A and B channels have all the crucial color information. These two channels are then classified using K-means clustering. K-means clustering treats each object as if they belong to some location in space. It finds partitions so that objects within a cluster are as close as possible to each other and far from an object that belongs to another cluster. However, K-means clustering requires the number of clusters and a distance metric to quantify proximity. In this research, City block is used rather than traditional Euclidian distance. The reason is for the high dimensional vectors the City block works better than Euclidean distance. Maximum iteration for clustering is 100 times, and in order to avoid local minima clustering is repeated 20 times. Clustering is an unsupervised method, so it is not possible to pre-define the best cluster based on features. However, every pixel in the image is labeled by the clustering results. For every object in the input image, K-means returns an index corresponding to a cluster. All the pixels in the image are labeled with a clustered index. In addition to that, pixels that do not belong to a cluster are set to 0(black). One of the most challenging tasks is to select the skin cluster.

To choose the skin cluster, the hue component has to be taken into account. Hue has obvious skin cluster quality, and the range of hue is relatively stable and thus can be regarded as a basic parameter for skin detection. However, human skin has a stable hue value. Hue value remains unchanged in relation to illumination intensity and the curvature of the skin's surface change. Another benefit of Hue is that it shows the highest contrast between skin and background. Besides, skin color has a very narrow range of hue in HSV color space. Hue range for skin is $[0^0 \text{ to } 60^0]$.

In order to further understand the range of hue for the creation of an efficient skin algorithm, an experiment was conducted on the Compaq dataset. Figure 3.14 shows a hue range of skin and non-skin pixels of Compaq dataset. The range of hue is almost $[0^0\text{-}60^0]$.

However same experiment was done on SFA dataset in Figure 3.15 and the result were similar. The hue range for skin is $[0^0-50^0]$. Only skin pixels of the IBTB dataset was used and shown in Figure 3.16. The hue range for only skin pixels only range between $[0^0-50^0]$.

FIGURE 3.14. Histogram of hue value for non-skin and skin of compaq dataset

FIGURE 3.15. Histogram of hue value for non-skin and skin of SFA dataset

FIGURE 3.16. Histogram of hue value for non-skin and skin of IBTB dataset

Based on this data, the mean hue value of all clusters is computed. Among them, the cluster with the least mean hue value is chosen for skin cluster. The reason that the smallest mean hue value cluster has the maximum probability of containing skin cluster. It can also be justified since the background of the cluster is black and black pixel has hue value zero. The mean hue value of skin cluster is only influenced by the non-black pixels of the cluster. An example is shown in Figure 3.17.

Original Image	Cluster 01 (Mean Hue)	Cluster 02 (Mean Hue)	Cluster 03 (Mean Hue)	Selected Cluster With Minimum Mean Hue
(0.0285)	(0.2624)

(0.458)	(0.013)

(0.2359)	(0.0446)	(0.0904)

(0.023)	(0.235)

FIGURE 3.17. Example to choose skin cluster

3.2. Region Growing

This clustering method performs well in comparison to other approaches, but it can be improved even further. This clustering method is an unsupervised method therefore, some skin pixels are missing due to some real-world condition. To solve this problem, a region growing approach is proposed to get the maximum possible skin pixel from an image [13]. As Table 3.1 shows, there are some skin pixel information missing after clustering. Region

growing algorithm can provide connected regions and have similar statistical characteristics with boundary information [53]. This whole process is described in Algorithm 3.

3.2.1. Region Growing Algorithm

	Input : Original Image
	Output: Skin region of Input Image
1	Input Image $I(r,c)$ where r = number of rows and c = number of columns
2	Call Selection of Seeding point
3	Get Index of seeding point (x,y)
4	Get Mean standard deviation σ
5	Locate the index of (x ,y) in Image(I)
6	Get index of four neighbors of Index(x,y)
7	a. Index of neighbour1($x-1,y$)
8	b. Index of neighbour2($x+1,y$)
9	c. Index of neighbour3($x,y-1$)
10	d. Index of neighbour4($x,y+1$)
11	Intensity of neighbors $\leftarrow mean(intensity(neighbour1,..neighbour4))$
12	Get the intensity value β at $I(x,y)$
13	foreach For each pixels in Image do
14	if Intensity Value $\leq$ Intensity value of neighbors then
15	Add neighbor pixels to region
16	else
17	Select new neighbors
18	end
19	end
20	Repeat previous step
21	Exit

Algorithm 3: Region Growing Algorithm

TABLE 3.1. Missing skin pixels from clustering

Original Image	Skin Detection by Clustering

For region growing two features are needed: initial seeding points and similarity metrics. The reason to select seeding points is to select pixels that are most obviously skin pixels. The performance of the region growing algorithm depends on this selection of seeding points. Further in this dissertation, it will be explained on how to ensure to get the obvious skin seeding points.

The selected cluster has either skin pixels or the background as black. Selected cluster is again clustered into two clusters. Specifically two clusters are chosen so that one cluster has the skin information and another will have background pixels. The algorithm described

Input : Skin image cluster

Output: Index of skin pixels

- 1 Perform K-means clustering for the Input image
- 2 Number of cluster = 2
- 3 Calculate the centroid value for each clusters C1, C2
- 4 Maximum Centroid =max(C1,C2)
- 5 Get Index of the Maximum Centroid
- 6 Get Mean standard deviation σ of skin areas only
- 7 Exit

Algorithm 4: Selection of Seeding Point

in Algorithm 4. K-means clustering was used for clustering. This is described in the Figure 3.18. Standard deviation of image intensity guides the region growing process. As the region is growing iteratively by comparing neighboring pixels, the difference between each pixels intensity value and the regions mean value becomes the measurement of similarity. If the difference value is smaller than the standard deviation the neighbor pixel is assigned to the region. Since standard deviation is used, it can recover the skin pixels that were previously missed by clustering. This region growing act as a compensation process to recover the skin pixels.

FIGURE 3.18. Calculation of skin seeding points

3.2.2. Region Growing Results

Figure 3.19 shows the recovery of missed pixels in region growing. In Figure 3.19(b), it is apparent that some of the skin pixels are definitely missing from the clustering approach. In Figure 3.19(c), most of the skin pixels are detected with the obvious skin seeding points and standard deviation of only skin pixel. Hence, the region growing compensates for missing

skin pixels. This technique can recover the missing skin pixels that are missed for various lighting or "real-life" challenges.

FIGURE 3.19. Recover skin pixels by region growing

3.3. Summary

In this chapter, proposed clustering and region based skin detection is explained. These methods take advantage of clustering methods with the homogeneity characteristic of human skin. Clustering methods require a number of clusters to be defined prior clustering. Images come in many varieties, and therefore a fixed number of clusters for all images is unrealistic. The proposed method as discussed throughout this chapter, count the number of significant objects in each image then calculate the area (by numbers of pixels). This process ensures that all the skin pixels are covered. An experimental result for three publicly available data set (Compaq, SFA, and HGR) was conducted to validate the number of the significant object detection technique. A new experiment shows that only the largest ten percent objects in an image contain most of the image pixels. This also applied for all types of images, including non-skin. This opens a new door for researchers working in the same domain. Then using the computation of the largest ten percent of the image's objects, a region growing algorithm is applied to recover skin pixels that are missing from clustering. In the next chapter, this proposed method will be applied to various datasets and the results will be compared with various established skin detection methods.

CHAPTER 4

EXPERIMENTAL EVALUATION

4.1. Dataset Description

Compaq dataset: For purposes of evaluation, The Compaq data set [37] was used. The data set is the most popular one for skin detection. It contains approximately 4600 images depicting skin and approximately 9000 images not containing skin. The entire database includes approximately 2 billion skin and non-skin pixels. Skin regions within 4,675 images containing skin have been segmented which in conjunction with non-skin images sums to 1 billion pixels.

This dataset was constructed from a larger image file collection downloaded from the internet. The set of images is divided into two groups: One group consists of images having human skin and the other consists of images having none as listed at Table 4.1. All images are color images. Individuals from different ethnic groups are represented. Also, the images depicting skin were taken under different angles, positions, and brightness conditions.

Percentages of images among three types are shown in Figure 4.1 are included. Examples from the Compaq data set are depicted in Table 4.2. In addition, all skin images are manually labeled using a software-assisted method. This tool segments skin regions by using connected component algorithm. At first, a threshold value, for which the value is a distance in RGB space, is used to select the skin region. None the less, multiple seeding points can be selected and with adaptation of the threshold, skin regions with complex shapes can be segmented.

Following with a manual step to label the skin, eyes, hair and open mouths are excluded. A binary mask (a monochrome image) for each image was constructed that depicts skin pixel on a black background of non-skin pixels. This was the truth data and it resides with the Compaq data set.

Obviously, applying the software-assisted manual approach is quite tedious and not a viable long-term solution to large-scale problems. Skin is sometimes obscured by different hair styles and it is almost impossible to have a perfectly precise segmentation. The authors

chose a conventional solution. Only identifiable skin pixels are marked and others discarded. Another important aspect of image collection from internet was color quantization. As image sources varied over various cameras, faxes, and other sources. These, of course, are the issues that make this a technical problem worthy of research.

TABLE 4.1. Compaq dataset details

Compaq Dataset			
Category	Sub-category	Number of Images	Descriptions
Non-Skin		8,964	These images do not contain any human skin.
Skin	Non-pornographic	3,841	These images contain individuals with different illuminations, complex backgrounds. People Included from different ethnicities, ages, and sex.
	Pornography images	828	Images are pornographic. Background of these images varies form complex indoor and outdoor background
	Swimwear images	151	Individuals wearing swimsuits. They have a significant amount of skin exposed.
Total		13,784

FIGURE 4.1. Percentage of skin image types in compaq dataset

TABLE 4.2. Example of Compaq dataset images

SFA Dataset: The SFA data set was constructed on FERET and AR databases [13]. Its purpose is the testing of classifiers based on skin color or texture features. It has been observed that use of color features for skin detection significantly improves accuracy. In construction, SFA used 876 images from FERET and 242 images from AR. Skin and non-skin samples were provided with their ground truth. Some examples of the SFA dataset are given in Table 4.4. Those data sets are comprised of images of various people including Caucasian, Asian, and African across a variety of ages and gender. The samples vary in size from 1 to 35x35 pixels. In the AR data set, all the images have a white background with little skin color variation. On the other hand, the FERET data set has high skin color and

various backgrounds.

In constructing the SFA data set, for each image three skin samples and five non-skin samples were taken. Each sample has a central pixel with obvious skin and non-skin properties. To generate the ground truth, every non-skin pixel was manually converted to black pixel. It is quite exceptional that, in building the data set, a validation step was also performed concurrently. This validation was performed using an Artificial Neural Network (ANN). Since this is supervised learning, training data is needed. This training process used scaled conjugate gradient learning as provided in the neural network tool box of Matlab. Lastly, all images were segmented pixel by pixel and ground truth images were binarized and used as a benchmarks for segmentation. The detailed statistics of the SFA dataset are given in Table 4.3.

TABLE 4.3. SFA dataset details

SFA Dataset	
	Properties
Total images	1118
Contents	Facial
Number of Total Pixels	1344 million (Approximate)
Number of Total Skin Pixels	386 million (Approximate)
Number of Total Non-Skin Pixels	958 million (Approximate)
Ground Truth Derived	Manually
Source	AR and FERET dataset
Lighting conditions	Indoor uncontrolled
Gender	Male/ Female
Skin Type	Various racial group
Color Space	RGB

TABLE 4.4. Examples of SFA dataset images

HGR dataset: The HGR dataset was built with a aim of assisting in training hand recognition classifiers. All images depict hand configurations from Polish and American Sign Language. It consists of 899 hand images from 12 individuals. The image dimensions vary from 174×131 to 640×480 . All images are captured under uncontrolled background and lighting conditions. The database was developed by [54]. The HGR dataset details are given in Table 4.5. Subjects in the dataset are male and female from various ethnic group. Examples of the images within it are listed in Table 4.6.

TABLE 4.5. HGR dataset details

HGR Dataset	
	Properties
Total images	899
Contents	Hand Gesture
Number of Total Pixels	190 million(Approximate)
Number of Total Skin Pixels	38.7 million(Approximate)
Number of Total Non-Skin Pixels	150 million(Approximate)
Ground Truth Derived	Manually
Source	HGR dataset
Lighting conditions	Indoor uncontrolled
Gender	Male/ Female
Skin Type	Various racial group
Color Space	RGB

TABLE 4.6. Example of HGR dataset images

4.1.1. Evaluation Metrics and Matrices

The choice of proper metrics for a classifier evolution is an important decision. This choice depends on application domain, final features selection, and obviously classification goal. Performance scores are used to make fair evaluations. In this dissertation, to achieve fair comparisons of the methods, several evaluation metrics, some expressed as matrices, were compared. There are impediments to comparing different methods using the same performance metrics. For example, every researcher uses his/her own evaluation matrices. In this dissertation, to compare the proposed approach with other classifiers, the metrics employed and published by authors of alternative approaches are adopted. It is worth noting that, evaluation metrics varied greatly. The most basic metrics used for this dissertation are Correct Detection Rate (CDR), False Positive Rate (FPR), False Negative Rate (FNR), True Negative Rate (TNR), F-measure, Precision, Recall, and Specificity. A confusion matrix is used to combine CDR, FPR, FNR, and TNR into one illustration.

Confusion Matrix An $n \times n$ confusion matrix is a table summarizing classifier performance. The rows and columns are labeled by class identifiers. Given a binary classification, skin/non-skin, the confusion matrix has two rows and two columns. This square matrix denotes actual class (the rows) vs. predicted class (the columns). The sum of the values in the matrix is equal to the number of samples in the experiment. That is, each experimental instance is aggregated in to only one element of the class label set $\{S, N\}$. The details are shown in Figure 4.2. A classifier maps from instances to the predicted class. However, some classifiers (known as multi-label classifiers) produce an estimation of instances in terms of probability. More often, however, a threshold is used to classify a single instance in to a single predicted class. In other words, most classifiers produce a distinct class label to predict the instance's class. To distinguish the actual class and the predicted class, usually a label set Positive, Negative for class prediction is used.

For a given classifier, there are four possible outcomes:

- If the experimental instance is positive and the predicted result is positive, it is counted as a true positive (TP)

		Prediction outcome	
		Skin	Non-Skin
Actual value	Skin	True Positive	False Negative
	Non-Skin	False Positive	True Negative

FIGURE 4.2. Confusion matrix of skin detection

- If the experimental instance is positive and the predicted result is negative, it is counted as a false negative (FN)
- If the experimental instance is negative and the predicted result is negative, it is counted as a true negative (TN)
- If the experimental instance is negative and the predicted result is positive, it is counted as a false positive (FP)

For the performance evaluation of the proposed classifier which assorts the skin and non-skin labels for each image pixel, the metrics below are used. The predicted class is compared with the provided ground truth. The details are given in Figure 4.3.

- **True Positive (TP):** If the ground truth pixel is skin and the classifier prediction is skin, then it is True Positive (TP)
- **False Negative (FN):** If the ground truth pixel is skin and the classifier prediction is non-skin, then it is False Negative (FN)
- **True Negative (TN):** If the ground truth pixel is non-skin and the classifier

FIGURE 4.3. Parameters of skin detection

detection is non-skin, then it is True Negative (TN)

- **False Positive (FP):** If the ground truth pixel is non-skin and the classifier detection is skin, then it is False Positive (FP)

Continuing with the definition of metrics, an example of experimental results in the form of a confusion matrix is provided in Figure 4.4. However, the confusion matrix is not itself a performance measure but most of the performance metrics are based on confusion it.

FIGURE 4.4. Confusion matrix example from an image

Accuracy/Correct Detection Rate (CDR) A frequently used performance measure known as accuracy, predictive accuracy, or correct detection rate (CDR). Accuracy is indicative of classifier error that can be evaluated over two or more classes.

$$(4.1) \quad Accuracy/CDR = \frac{TP + TN}{TP + TN + FP + FN}$$

In skin detection, accuracy indicates the fraction of pixels correctly classified as skin or non-skin based on ground truth.

False Positive Rate (FPR) indicates the fraction of pixels that are non-skin but classified as skin pixels. It is calculated as the total number of FP pixels divided by the number of pixels for which the truth data specifies non-skin. Lower values of FPR indicate good classifier performance. It is also known as type I error (also “consumer error”).

$$(4.2) \quad FalsePositiveRate(FPR) = \frac{FP}{FP + TN}$$

False Negative Rate (FNR) is the fraction of pixels ground-truthed as skin pixels but classified as non-skin. This is also known as type II error (also “producer error”).

$$(4.3) \quad FalseNegativeRate(FNR) = \frac{FN}{FN + TP}$$

Specificity/True Negative Rate (TNR) is the likelihood that a non-skin pixel is classified as non-skin. Specificity measures a classifier’s performance to correctly detect members of the negative class (non-skin). A higher specificity will detect more non-skin pixels. It focuses on the negative class.

$$(4.4) \quad Specificity/TrueNegativeRate(TNR) = \frac{TN}{TN + FP}$$

Sensitivity/Recall/Hit Rate/True Positive Rate (TPR) indicates a classifier's reliability in detecting skin pixels. If a classifier has high sensitivity, it will successfully predict 80 percent of the pixels that are skin pixels. But the set classified as skin pixels will contain 20 percent that are not skin pixels. It focuses on the positive class.

$$(4.5) \quad \text{Sensitivity/Recall/HitRate/TruePositiveRate}(TPR) = \frac{TP}{TP + FN}$$

Precision indicates the number of true positive pixels detected by the classifier. In this case, precision is the proportion Of skin pixels correctly predicted by the classifier. The metric is the number of positive predictions divided by total number of positive samples. It is also known as the positive predictive value (PPV). Precision measures the exactness of the classifier.

$$(4.6) \quad \text{Precision} = \frac{TP}{TP + FP}$$

F₁-measure is generally considered the best metric for the comparison of one classifier with a second. Basically, the F₁-measure is the harmonic mean of precision and recall¹. The F₁-measure quantifies the balance between precision and recall. Precision and recall are inversely proportional to each other and the score assists in reconciling a contradiction that may appear between them. The score takes both false positives and false negatives into account. It expresses the preciseness of a classifier. In other words, it is proportional to the number of instances detected correctly which is related to the classifier's robustness. However, a high precision and low recall indicates that the classifier is very accurate but at the cost of a large miss. Intuitively it is not easy to understand accuracy and the F₁-measure is usually more useful, especially if there is an uneven class distribution. Mathematically it is defined as

¹The harmonic mean of $x_1, x_2, \dots, x_n$ is $H = \frac{n}{x_1^{-1} + x_2^{-1} + \dots + x_n^{-1}}$ which has the special case that if $n = 2$ then $H = \frac{2x_1 \cdot x_2}{x_1 + x_2}$.

$$(4.7) \quad F_1 - measure = \frac{2}{Precision^{-1} + Recall^{-1}}$$

$$(4.8) \quad = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Note that it is the harmonic mean of a sequence of two values – precision and recall.

4.1.2. Pixel based Skin Detection for Compaq Dataset

Skin detection is frequently the pre-processing method of choice for computer vision applications that focus on human activity, identity, or appearance. Researchers have considered many features for skin detection. Usually, color-based skin detection methods are applied. A number of color spaces are available and distinguishing skin color from the background is often as simple as implementing a threshold operator. If the skin is to be so classified, there are two major approaches – pixel based and region based.

In pixel-based detection, each image pixel is classified as skin or not using a statistical method together with a single color space or the fusion of several spaces. That is detection uses purely local properties. On the other hand, region-based detection uses quantitative measures over regions. Small, homogeneous regions are sometimes called super pixels.

In this dissertation, both approaches are compared with the proposed method. Three well known datasets are used for evaluation. These data sets have been used by a number of researchers and their results are published. Those results are compared with the same metrics as the proposed method. This ensures the integrity of the comparisons and fair evaluation.

Non-parametric methods estimate the skin class without assuming a specific probability density function(pdf) or knowledge of population parameters. It is the lack of lack of this knowledge generally leads to the need for more data and more training time than that needed for methods that assume parametric information. Examples of non-parametric learning methods are decision trees, support vector machines, and k-nearest neighbor clustering. Notably, no non-parametric methods were identified relevant to the topic of this dissertation.

TABLE 4.7. List of methods compared for Compaq dataset

Categories	Methods	Research Paper
Parametric	Bayesian	M. Jones, J. Rehg, Statistical color models with application to skin detection, IJCV 46 (1) (2002) 8196.
	Adaptive Bayesian	Hung-Ming Sun, Skin detection for single images using dynamic skin color modeling, Pattern Recognition 43 (4) (2010) 14131420.
	Bayesian	Hung-Ming Sun, Skin detection for single images using dynamic skin color modeling, Pattern Recognition 43 (4) (2010) 14131420
	Mixture of Gaussian	M. Jones, J. Rehg, Statistical color models with application to skin detection, IJCV 46 (1) (2002) 8196.
	Elliptical Model	J.Y. Lee, S.I. Yoo, An elliptical boundary model for skin color detection, Proceedings of the International Conference on Imaging Science, Systems, and Technology, 2002
	SOM	D. Brown, I. Craw, J. Lewthwaite, A SOM based approach to skin detection with application in real time systems, in: BMVC, 2001.
Fusion	RGB ETM	J. Kovac, P. Peer, F. Solina, 2D Versus 3D Colour Space Face Detection, 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, 2003, pp. 449454
	NNDST	T. Denoeux, A neural network classifier based on DempsterShafer theory, IEEE Transactions on Systems, Man and Cybernetics A 30 (2) (2000) 131150
	Majority	Francesca Gasparini, Silvia Corchs, Raimondo Schettini, Recall or precision-oriented strategies for binary classification of skin pixels, Journal of Electronic Imaging 17 (2) (2008).
	Product	
	Threshold	
	Platzer	C. Platzer, M. Stuetz, M. Lindorfer, "Skin sheriff: A machine learning solution for detecting explicit images", Proc. 2nd Int. Workshop Security Forensics

Parametric methods include the Bayesian decision rule which is very effective in statistical pattern classification. Other parametric methods include single Gaussian, mixture-of-Gaussian (MoG) and elliptical boundary models. All three require training samples for model generation. Training generally requires only a small number of samples. Expectation maximization is used for estimating approximate parameters such as mean, median or covariances. Methods which use the Gaussian model assume that skin pixel values have a color distribution represented by a Gaussian pdf. But skin color boundaries are seldom specified accurately. However, mixture of Gaussian is a fusion of two or more Gaussian components and can handle multiple luminance levels. However, luminance levels vary sufficiently in real world and Gaussian/MoG methods are less accurate than generally desired.

Recall that parametric methods use approximated parameters, thus detecting skin cluster boundaries inexactly. The training input for a Bayesian classifier is a discretely represented histogram, not a continuous pdf. Ultimately, the solution is in the form of thresholds distinguishing positive and negative classes. Nevertheless, the Bayesian classifier exhibits higher detection rates compared to other classifiers. To compare the results the methods listed

TABLE 4.8. Comparison of accuracy and false positive rate of statistical methods

Categories	Methods	Accuracy	False Positive Rate (FPR)
Parametric	Bayesian 1	90	14.2
	Adaptive Bayesian	69	5
	Bayesian 2	74.5	5
	Mixture of Gaussian	90	15.5
	Elliptical Model-xyz	90	20.9
	Elliptical Model-YCbCr	90	33.33
	Elliptical model-YIQ	90	30
	SOM	78	32
Fusion	RGB ETM	87.01	21.88
	NNDST	90.01	19.07
	Majority	82	18
	Product	32	4
	Thresholds	89	21
	Platzer	82.3	11.4
	Proposed	92.04	5.97

in Table 4.7. The accuracy and FPR comparison result for various statistical methods are depicted in Table 4.8.

Figure 4.5 shows the accuracy comparisons of of all parametric methods with the proposed method. The proposed method has 2.26 % more accuracy than its nearest competitor Bayesian1 and 23% and 17.54% more accuracy than Adaptive Bayesian and Bayesian2 methods, respectively.

FIGURE 4.5. Accuracy comparisons of parametric methods

Elliptical models where three different color spaces (xyz, YCbCr and YIQ) were used. However, it is quite remarkable that a popular artificial neural network (ANN) uses a self-organizing map (SOM) for skin and non-skin distributions. Its performance lags that of other methods. The reason is, SOMs are an unsupervised method, they depend on training data for a better classification result. Many non-skin pixels were not included in training data and that degrades performance.

Explicit threshold based methods (ETM) uses predefined thresholds. ETMs are easy to implement. They do detect skin at the boundaries of skin clusters. That simplifies the requirement of training. However, when the boundary is applied to individual pixels, it is biased toward false positive outcomes if the thresholds are large for all color variations. In addition to that, all the color bands possess the same weight, so mis-classification decisions propagate. Lastly, threshold-based methods are not sufficiently flexible to capture all variations.

Fusion-based methods get attention because they can reduce uncertainty of features. Some authors fuse at the decision level. Alternatively, some authors create new features by fusing the signal-level features - the chrominance components. For example, Gasparin, et al. [27] use different feature fusion rules - the sum rule, the product rule, the majority vote, and threshold-based methods. Their hypothesis is the combination of bands yields better performance. Yet the authors conclude most of the fusion rules performed poorly and were not sufficiently robust. A more robust and flexible method is needed.

Figure 4.6 depicts accuracy comparisons with the proposed method to existing fusion-based methods. Among the surveyed methods, NNDST [20] showed the most competitive performance (2.03% less than the proposed method). NNDST is a neural network based classifier employing Dempster-Shafer theory (DST). This method is able to overcome uncertain and incomplete training data. However, it still requires a large amount of training data. As for other methods, product rules perform the worst. An explanation is: Instead of combining various threshold methods, it only classify pixels as skin that are detected by all thresholds.

FIGURE 4.6. Accuracy comparisons of fusion based methods

Discussion of results. It is obvious that the proposed skin detection method outperforms existing approaches for the Compaq dataset. Comparisons using other statistical methods, the proposed method continues to outperform. (This is discussed in more detail below.) The performance is due to the fact that the proposed approach takes advantage of clustering methods and the LAB color space. Overall accuracy comparisons with all statistical methods are shown in Figure 4.7.

False Positive Rate (FPR) indicates the fraction of non-skin pixels detected as skin as defined above. FPR does reflect the proportion of each individual negative class (hair, clothes, etc.) incorrectly classified as a positive class. A low false positive rate is preferred to a higher one.

Figure 4.8 shows that few of the statistical methods (adaptive Bayesian, Bayesian2, and product) have a lower FPR than the proposed one. However, the benefit comes with a trade off in accuracy. False positive rates alone are but a minor evaluation criterion for measuring classification efficacy. Comparisons of accuracy combined with FPR are given in Figure 4.9.

Performance of the classifier in terms of accuracy and FPR can be illustrated by the gap between accuracy and FPR. In other words, if a method has lower FPR and higher accuracy then it classifies both positive and negative instances well. A few of the statistical

FIGURE 4.7. Accuracy comparisons of all statistical based methods

FIGURE 4.8. False positive rate(FPR) comparisons of all statistical based methods

approaches have high accuracy – Bayesian1 (90%) with a FPR rate of 14.2%, the elliptical model with all color spaces has 90% accuracy with a high FPR rate of 20.9%, 33.33% and 30%, respectively. Among the surveyed methods NNDST has 90.01% accuracy with moderate FPR of 19.07%. On the other hand, adaptive Bayesian, Bayesian2, and product have minimum FPRs but their accuracy suffers. All are less than 80%. Lastly, considering simultaneously both metrics, the proposed method has the highest accuracy at 92.04% with the second lowest FPR at 5.97%. So, the proposed method outperforms other surveyed

FIGURE 4.9. Accuracy and FPR for all statistical methods

statistical methods in terms of combined accuracy and FPR.

The Figure 4.10 depicts an example of skin detection on a complex background for several statistical methods. The Bayesian classifier detects skin better but also misclassifies some background pixels. The MoG classifier exhibits poorer results than other methods. The RGB ETM performs better compared with each of the others. However, the proposed method, detects skin pixels better than any other method as shown in Figure 4.10(f).

4.1.3. Color based Skin Detection for Compaq Dataset

Several of the trichromate color spaces are used to classify skin from images by means of fixed thresholds. Some of the methods employ pre-processing. Christin [57] reported skin detection results for both single color spaces and fused color spaces. All exhibited high recall and false positive rates but having low precision rates. The fusion of two or more color spaces is superior but still there is room for improvement. The results are recorded in Table 4.9.

As Table 4.9 indicates, threshold-based methods have a higher recall than the proposed one. RGB-1D and RGB have 18.25% and 20.15% better recall. However, high recall may indicate the classifier assigns labels to many pixels as skin including most actual skin pixels. Meanwhile, many others assigned as “skin” are not skin pixels. The latter results in a precision value that is low compared with the proposed method. High precision indicates that the classifier is more accurate considering only actual “skin” pixels and correctly labels

(a) Original Images (from the Compaq database).

(b) Detection by the Bayesian classifier.

(c) Detection by the MoG classifier.

(d) Detection by the RGB ETM.

(e) Detection by the NNDST.

(f) Detection by the Proposed Method

FIGURE 4.10. Comparison of skin detection(proposed vs statistical methods)

TABLE 4.9. Single color space based skin detection [57]

Method	Recall	Precision	FPR	Accuracy	F-measure
RGB-1D	0.91	0.204	0.295	0.721	0.333
RGB	0.929	0.197	0.315	0.704	0.325
HS	0.78	0.2	0.259	0.744	0.319
YUV-YIQ	0.874	0.191	0.308	0.706	0.313
LCCS	0.641	0.188	0.229	0.761	0.291
HSV-1	0.855	0.163	0.364	0.653	0.274
YCbCr-1	0.776	0.163	0.331	0.678	0.269
HSV-2	0.926	0.144	0.459	0.571	0.249
MDL	0.988	0.124	0.582	0.461	0.22
HSI	0.995	0.112	0.653	0.397	0.202
TSL	0.326	0.128	0.184	0.779	0.184
YCbCr-2	0.526	0.1	0.393	0.601	0.168
YUV	0.564	0.087	0.492	0.512	0.15
Proposed	0.7275	0.625	0.059	0.92	0.64

High recall, low precision **Low recall, high precision**

FIGURE 4.11. High and low recall and precision

most such pixels. That is, the pixels labeled as skin are skin while it may mis-label a lot of actual skin pixels. Precision and recall have a very interesting inverse relationship. It depends on the particular aim of a classifier as pertains to balancing the cost of errors on the positive side vs. the cost of errors on the negative side. In this work, each error is considered to have the same cost. Figure 4.11 has been adopted from <https://medium.com/@klintcho/explaining-precision-and-recall-c770eb9c69e9>. Free License.

For many reasons, precision and recall do not indicate the performance of a classifier perfectly. Most approaches balance the costs between precision and recall based on minimizing the total costs. Both high precision and high recall is desirable but not realistic. Remember that for a single, not fused, color space, each of the high recall methods suffers from low precision. As Figure 4.12 shows, there is a trade off between precision and recall for the proposed method. The gap between precision and recall indicates the balance between these two evaluation metrics. A large gap signifies better balance. The proposed method has a good balance, i.e. large gap, and outperforms other methods.

However, for a fairer comparison, another evaluation metric, namely the F-measure, is often used. The F-measure is the harmonic mean of the pair of values precision and recall. The F-measure is used for classifier performance and it balances precision and recall when the cost of an error is the same for both classes. It is a more realistic measure than precision and recall individually. The F-measure is also more useful than accuracy when given uneven class distributions.

FIGURE 4.12. Recall, Precision, FPR, Accuracy and F-measure of single color space

FIGURE 4.13. Overall performance of single color space

In summary, Figure 4.13 shows the overall performance as box plots. Each box in the figure is sourced from one column in Table 4.9. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points omitting outliers. Outliers are plotted individually using the '+' symbol. In terms of recall, the median was over 0.8 which is higher than the proposed classifier of 0.72. However, the median for precision is below 0.2 and the proposed method is an outlier with 0.625. For false positive rate (FPR) the median value is around 0.3. For the proposed method it is 0.059 which is better than any of the others. In addition, the accuracy median value is below 0.7 but the proposed method has an accuracy of 0.92. Lastly, for the F-measure the median is around 0.3 while the proposed method is an outlier with the impressive performance of 0.64. So, for each of the performance metrics except recall, the proposed method outperforms existing single color space based skin classification.

Combined color spaces show improvements over single color methods but the proposed

method still outperforms them. The evaluations are given in Table 4.10. To compare the options, recall, precision, FPR, accuracy, and F-measure metrics are used.

TABLE 4.10. Combined color space based skin detection [57]

Method	Recall	Precision	FPR	Accuracy	F-measure
RGB+HS	0.757	0.263	0.176	0.819	0.39
RGB1D + RGB	0.87	0.241	0.227	0.78	0.378
RGB+HSV2	0.893	0.239	0.236	0.774	0.377
RGB+HSV1	0.813	0.236	0.219	0.784	0.366
RGB+YUV-YIQ	0.854	0.23	0.237	0.77	0.363
RGB+HCBcCr	0.61	0.233	0.166	0.816	0.338
RGB+YCbCr1	0.73	0.21	0.228	0.769	0.327
RGB1D+ YCbCr1	0.74	0.204	0.24	0.758	0.319
RGB1D +HCBcCr	0.639	0.199	0.214	0.775	0.303
RGB+YCbCr2	0.48	0.188	0.172	0.802	0.27
RGB1D +YCbCr2	0.491	0.184	0.181	0.794	0.267
Proposed	0.7275	0.625	0.059	0.92	0.64

In the case of precision, the proposed method performs better than any of the tested combination of color spaces. The reason is that functionally combining color information actually increases the difficulty in distinguishing between skin and non-skin pixels. That is to be expected because if precision is higher then the FPR will be less for the classifier. Nevertheless, mixing RGB and HS performs very nearly the same as the proposed method which is 10% lower.

Accuracy is the metric most often used for comparing classifier performance. For combinations of color space, accuracy increases because it rules of combination can take advantage of the additional information. A graphical comparison is given in Figure 4.14. However, mixing colors also causes confusion. Nonetheless, the proposed method uses only one color space, omitting the illumination component. It provides better accuracy compared to each of the others.

For FPR, as shown in Figure 4.15, only 5.9% are mis-classified by the proposed method. Other tested methods have a higher FPR rate ranging from 17%-23%. Because a lower FPR is desirable, the proposed method performs best among the selected alternatives.

The tested color-fused methods have lower recall values compared with the tested single

FIGURE 4.14. Accuracy of combined color space

FIGURE 4.15. FPR of combined color spaces

color space methods as shown in Figure 4.16. Yet, as to be expected, the precision is higher across the board. Two options, the combination of RGB with HSV and YIQ, have higher recall but much lower precision than the proposed method. The difference between proposed method precision and the others is almost 40% as shown in Figure 4.17.

The F-measure, being a balance between precision and recall, provides a good indication of classifier performance. As the black line indicates in Figure 4.18, the F-measure is highest for the proposed method.

FIGURE 4.16. Recall of combined color space

FIGURE 4.17. Precision of combined color space

Overall, fused color space based skin detection performs better than the single color space as evident in Figure 4.19. However, the median value of recall of the former is just below 0.8 and proposed method reaches just 0.72. That is below the median. Compensating is the impressive precision of 0.62 where as the median precision value is below 0.3 for fused color space. In addition the FPR median value is little more than 0.2 but the proposed method has 0.059. That is better by a factor of three and indicated as an outlier, denoted as "+" in Figure 4.19. For accuracy it has the same trend. The proposed method is almost 1.5 times better than the median value over all fused color options. Lastly, The F-measure

FIGURE 4.18. Precision, Recall, and F-measure of combined color space

FIGURE 4.19. Overall performance of combined color space

for the proposed method is an outlier, indicating better performance of the classifier. Remember, among these performance metrics, F-measure is the most effective in comparing performances.

To compare with not only the explicit threshold-based method but other methods as well, the result recorded at [42] is used. This paper provides an evaluation of color-based

skin detection for six color spaces – HSI, RGB, IHLS, YCbCr, CIELAB, and normalized RGB (nRGB) and eight skin color modeling approaches – AdaBoost, Bayesian network, J48, Multilayer Perceptron (MLP), Naive Bayesian, RBF network, SVM, and a histogram-based approach as listed in Table 4.11.

TABLE 4.11. F-measure of different methods for various color spaces [42]

Methods	IHLS	HSI	RGB	nRGB	YCbCr	CIELAB
AdaBoost	0.250	0.257	0.259	0.213	0.280	0.160
BayesNet	0.413	0.400	0.292	0.332	0.391	0.413
J48	0.540	0.588	0.550	0.593	0.560	0.567
MLP	0.549	0.481	0.602	0.527	0.571	0.534
NaiveBayes	0.284	0.220	0.292	0.389	0.227	0.423
RBF	0.418	0.186	0.351	0.327	0.333	0.501
SVM	0.462	0.438	0.312	0.204	0.227	0.418
Hist.	0.370	0.352	0.401	0.333	0.347	0.354
Proposed	0.64	0.64	0.64	0.64	0.64	0.64

The publicly available dataset, Compaq, is used to evaluate the F-measure. The authors intent is to create a base line that can be used to test future classification algorithms. The skin modeling approach AdaBoost uses a multi-class machine to solve the two-class problem. Bayesian networks use random variables and conditional independence of random variables. J48 builds a binary decision tree from a training dataset using information entropy. The performance is affected by the confidence factor and number of instances per leaf.

A multilayer perceptron (MLP) is configured by a sequence of layers composed of nodes. Each pair of adjacent layers is fully connected via weighted edges. Trained using labeled samples, the summed output error is used, through back propagation, to adjust the weights on the edges. The naive Bayesian method uses a posteriori probability to decide each test instance. Similar to a MLP, RBF network (RBF) takes radial basis function for activation while the MLP uses a hysteresis curve. SVM searches for a manifold that creates an inter-class separation. The goal is to establish the maximal margin between the manifold and instances nearest it. Lastly, the histogram method requires building a skin and non-skin pdf model and uses conditional probability for classification.

The comparisons using the F-measure is given in Figure 4.20. There is another phenom-

enon worth noting in Figure 4.20: CIELAB (a.k.a. LAB), utilized in the proposed method, has a better F-measure than each of the other five color spaces reported in the figure. The nearest competitor method and color space is RGB using an MLP classifier having an F-measure of 0.602 which is 3.8% less than the proposed method. The green bar indicates the proposed method's performance. However, it is also important to notice that, various classifiers using the same color space as the proposed method also have a lower F-measure. J48 has the closest performance with a difference of 7.3%.

FIGURE 4.20. F-measure for six color spaces of eight skin classifiers

4.1.4. Region based Skin Detection for Compaq Dataset

Pixel-based skin detection methods are dominate in terms of deployment in the field. Pixel-based methods label each pixel as skin or non-skin based only on pixel value. But, skin can be best be described as regions and, theoretically, region-based methods should perform better. Rudra [58] proposed a segmentation technique that is based on super pixels which group similarly colored pixels together. Instead of individual pixels these regions are classified as skin or non-skin. This technique outperforms others and this result is further improved when labeling rules are driven by conditional random fields (CRF). While

TABLE 4.12. Region based skin detection results for Compaq dataset [58]

Method	Accuracy	FPR
Rudra P K Poudel (SuperPixel Only)	91.44	13.73
Rudra P K Poudel (SuperPixel +CRF)	91.77	13.12
Jedynak et al., 03	82.9	10
Proposed Method	92.04	5.97

region based skin detection shows better performance than pixel-based skin detection as illustrated in Table 4.12, the proposed method still has a better performance: Accuracy is 0.6% higher and for false positive rate almost 3% lower. Some experimental results are given in Figure 4.21.

FIGURE 4.21. Region-based skin detection

Pixel-based methods allow some non-skin pixels to be labeled as skin but region-based with CRF omits much non-skin but fails to detect all skin pixels. Lastly, the proposed method successfully detects skin and non-skin pixels even more effectively.

4.1.5. Skin Detection for SFA Dataset

Many research papers base algorithm evaluation on the Compaq dataset. Some researchers focus on the skin detection in the absence of a complex background. This work is often evaluated via the SFA dataset. Pooya Tavallali [81] in a paper published in 2019 evaluated results using the dataset. To compare this recent algorithm with the proposed method, an experiment was performed using the dataset and comparing the result with that recorded in [81]. Tavallali [81] uses statistical information of pixels and their neighbors. To reduce the complexity, a cascade classifier based on the AdaBoost algorithm is described in the paper. Two edge detection methods – Canny and Sobel – were also used to increase classifier performance. The results from the following research papers listed at Table 4.13 are used for comparing Tavallali’s method with the proposed method.

Pooya Tavallali [81] published skin detection results in terms of true positive (recall), false negative, true negative, false positive, accuracy, precision and F-measure as shown in Table 4.14. It is important to note that the F-measure has a coefficient of 0.5 (generally written as $F_{0.5}$) rather than 1 for all experimental results ². The results are shown in Figure 4.22. For true negative rate, AdaBoost Sobel has the best performance. That means, it can detect non-skin images better than others. In terms of false positive, the proposed method has the lowest with a slightly better performance than AdaBoost Sobel. Moreover, for accuracy, the proposed method is 5.3% higher than the other methods. AdaBoost Canny has 0.4% lower precision rate.

² $\overline{F}_\beta = (1 + \beta^2) \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}$

TABLE 4.13. List of methods compared for SFA dataset

Methods	Research paper
Adaboost (Baseline)	Tavallali, Pooya, Mehran Yazdi, and Mohammad Reza Khosravi. "Robust cascaded skin detector based on AdaBoost." Multimedia Tools and Applications 78, no. 2 (2019): 2599-2620.
AdaBoost (Canny)	
AdaBoost (Sobel)	
RGB	Kovac J, Peer P, Solina F (2003) Human skin color clustering for face detection. EUROCON 2003, Comput Tool IEEE Region 8 2:144148
HSV	Wang Y, Yuan B (2001) A novel approach for human face detection from color images under complex background. Pattern Recogn 34(10):19831992
ER/GH	Gomez G, Sanchez M, Sucar L (2002) On selecting an appropriate color space for skin detection. Mexican Int Conf Artif Intell (MICAI): 6978
CMYK(base)	Sawicki DJ, Miziolek W (2015) Human color skin detection in CMYK color space. J Image Process, IET 9(9):751757
CMYK(ROC)	
CMYK(blue)	
CMYK(Cascade)	
YCbCr	Chai D, Ngan KN (1999) Face segmentation using skin-color map in videophone applications. IEEE Trans Circ Syst Video Technol 9(4):551564

TABLE 4.14. SFA dataset experimental results [81]

Method	TP (Recall)	FN	TN	FP	Accuracy	Precision	F(0.5) score
Adaboost Baseline	0.6672	0.3334	0.9478	0.0522	0.8751	0.9274	0.8602
AdaBoost Canny	0.4712	0.5288	0.9645	0.0355	0.8222	0.9299	0.7783
AdaBoost Sobel	0.2367	0.7633	0.9722	0.0278	0.7519	0.8948	0.575
RGB	0.511	0.489	0.857	0.143	0.684	0.7813	0.7065
HSV	0.594	0.406	0.687	0.313	0.641	0.6549	0.6417
ER/GH	0.339	0.661	0.917	0.083	0.628	0.8033	0.6305
CMYK(base)	0.75	0.25	0.753	0.247	0.751	0.7522	0.7517
CMYK(ROC)	0.75	0.25	0.759	0.241	0.754	0.7568	0.7554
CMYK(blue)	0.913	0.086	0.48	0.52	0.697	0.6371	0.678
CMYK(Cascade)	0.75	0.25	0.75	0.25	0.75	0.75	0.75
YCbCr	0.499	0.501	0.861	0.139	0.68	0.7821	0.5379
Proposed	0.829	0.17	0.971	0.028	0.929	0.9342	0.8692

FIGURE 4.22. F-measure for six color spaces of eight skin classifiers

FIGURE 4.23. Overall performance for skin detection using SFA dataset [81]

The box plots shown in Figure 4.23 show the overall performance of various methods using the SFA dataset. The median recall is close to 0.65 and 75th percentile are approximately 0.75. There are five methods that have higher recall (CMYK(base), CMYK(ROC), CMYK(blue), CMYK(cascade), and the proposed method). Only four methods reside below the 25th percentile.

False negative has a median below 0.4. The false negative of the proposed method is 0.17 which is second lowest. Only CMYK(blue) has a better false negative rate. The false negative rate represents actual skin pixels classified as non-skin. The reason CMYK(blue) performs better is the image backgrounds are mostly uniform.

True negative detects actual non-skin as non-skin. The median is approximately 0.85 and 75th percentile is approximately 0.95, with the highest true negative yielded by the proposed method – 97.1.

Median false positive is approximately 0.15 and the proposed method has 0.028. The best false positive rate is exhibited by AdaBoost Sobel which is only marginally 1% higher.

For accuracy, the proposed method outperforms the other algorithms and is 5.39% higher than the closest alternative, AdaBoost Baseline. Precision of AdaBoost Baseline and AdaBoost Canny exhibited promising results of around 0.92 which is higher than the median

precision value of 0.8. But the proposed algorithm has the highest precision – 0.9342. As mentioned, the values of recall and precision requires a trade off and F-measure gives clearer view of the classifier performance. The median value of F-measure is approximately 0.75. The two that fall above the 75th percentile are AdaBoost Baseline and proposed algorithm. Lastly, the proposed method has the highest F-measure at 0.8692. So, it is quite obvious that there exist methods that perform better with respect to some performance metrics but it is unusual that one method performs better in almost all performance matrices. The proposed algorithm outperforms with respect to all metrics except recall.

A recent paper by Rodrigo [25] hypothesises that threshold-based rules can be reversed and skin pixels are not isolated. The authors used an original hypothesis, the reverse of the original hypothesis, a combination of both, and consideration of neighbors. Results were reported in terms of precision, recall, and F-measure and are shown in Table 4.15.

Similar evaluation metrics were used by Brancati [7]. The authors used clustering based on optimized CbCr skin color, and original CbCr with an MLP-based classifier that uses four neurons per hidden layer joint with a Bayesian classifier and, most importantly, threshold-based method using the LAB color space Qiang [61]. The outcome is given in the latter rows of Table 4.15.

TABLE 4.15. Experimental comparisons for SFA dataset [25]

Hypothesis	Precision	Recall	F-measure
Original	0.8636	0.4214	0.5664
Reverse	0.8563	0.773	0.8125
Combined	0.9288	0.3958	0.5551
Neighbors	0.9176	0.5111	0.6565
Optimized-CbCr	0.39	0.89	0.5
CbCr	0.21	0.99	0.34
MLP	0.4	0.79	0.49
Nave Bayes	0.34	0.86	0.46
Lab	0.37	0.61	0.43
Proposed Method	0.9342	0.829	0.8692

Figure 4.24 shows a comparative evaluation of precision, recall and F-measure for the SFA Dataset [25]. The proposed algorithm has both the highest precision and the highest

F-measure. In terms of recall Optimized-CbCr outperforms the others. The overall perfor-

FIGURE 4.24. Precision, Recall and F-measure of SFA dataset [25]

mance companions is shown in box plot Figure 4.25.

For precision, the median value is 0.65, which is somewhat low. Despite a few methods have high precision, most are low. The proposed approach has the highest precision with a value of 0.9342. Recall has a median of 0.8 and proposed algorithm is the one with that

FIGURE 4.25. Overall performance of SFA dataset [25]

median. The F-measure has a median of over 0.5 and the proposed method has a F-measure of 0.8692 which is 5.62% higher than the closest alternative, the reverse hypothesis method.

Overall, the proposed method was compared with the most recent skin detection algorithms using the Compaq data set, it outperforms all tested methods in terms of F-measure which is a robust and reliable performance metric.

4.1.6. Skin Detection for HGR Dataset

Bran [8] presents a pixel-based skin detection using the YCbCr color space. This approach takes advantage of combination of chrominance values for YCb and YCr color subspaces for skin detection. However, the combination depends on the generated skin cluster. The experimental results were compared with six well-known pixel-based methods for publicly available HGR dataset. Four methods tested on variations of the YCbCr color space were: [32] with two different variations; YCbCr space and CbCr subspace [14]; Basilio [5] using CbCr subspace. Two methods tested on other color spaces were: Kovac et al [44] with evaluation in the RGB space; Sobottka and Pitas [73] evaluated in the HSV color space. All these published results have been compared with the proposed method.

TABLE 4.16. Experimental results for HGR dataset [32, 14, 5, 44, 73]

Methods	Precision	Recall	Specificity	F-Measure
Hsu et al(YCbCr space)	0.1825	0.9822	0.1873	0.3078
Hsu et al(CbCr space)	0.7025	0.4958	0.9488	0.5814
Chai and Ngan	0.6702	0.982	0.7543	0.7967
Basilio et al.	0.6955	0.9176	0.7912	0.7913
Kovac et al.	0.4753	0.7551	0.8806	0.5834
Sobottka and Pitas	0.5969	0.8353	0.7691	0.6963
Original-Hypothesis	0.8938	0.7664	0.9274	0.8252
Reverse-Hypothesis	0.7929	0.8429	0.8337	0.8171
Combined-Hypothesis	0.8994	0.6952	0.939	0.7843
Neighbors based	0.8818	0.7935	0.9211	0.8353
Proposed Method	0.896	0.888	0.9531	0.8919

Table 4.16 shows an interesting pattern. A new performance metric, namely specificity, is used. Specificity is another name for true negative rate (TNR) and indicates the proportion of true negatives labeled by the classifier. However, the original-hypothesis based approach

and combined-hypothesis approach perform better when measured by precision. These two methods use the YCbCr color space and the proposed algorithm uses LAB color space. In addition for recall, Hsu's first method using YCbCr and CbCr and Basilio's using CbCr have better recall than the proposed method. However, the balance between precision and recall, i.e. F-measure, is better for the proposed algorithm for each of these five. For instance, the original-hypothesis and combined-hypothesis methods have better precision but have a mediocre recall performance (0.7664 and 0.6952, respectively) which is 12.16% and 19.28% lower than the proposed algorithm, respectively. The results shown in Figure 4.26 demonstrate that F-Measure values are greatest for the proposed method. Lastly, the specificity is highest for the proposed method which indicates that this algorithm detects non-skin pixels better than the others.

FIGURE 4.26. Precision, Recall and F-measure for HGR dataset [32, 14, 5, 44, 73]

The overall performance is shown in Figure 4.27. The median precision is around 0.7 and proposed method has 0.89 which is the highest. However, median recall is around 0.85 (lower than the proposed method) and Hsu's CbCr method is the outlier on the low side. Median specificity is over 0.8 and the highest specificity was earned by the proposed method. Hsu's first method (YCbCr color space) showed worst performance for specificity and is indicated as an outlier. For F-measure, the highest value of 0.8692 is posted by the proposed method.

FIGURE 4.27. Overall performance of HGR dataset [32, 14, 5, 44, 73]

A recent paper by Rahmat [62] combines the luminance channels of three prominent color spaces HSV, YCbCr and normalized RGB. The results are listed at Table 4.17. A number of fusion and selection rules led to seven options, some including luminance and some omitting it. Given a selection/fusion option, clustering was the chief labeling operation. The evaluation results are given in Table 4.17.

In Figure 4.28, the accuracy of each option is shown as a blue bar and the proposed

TABLE 4.17. HGR dataset results published by [62] for skin detection

Method	Accuracy	FPR	FNR
HSV	93.36	1.9	6.64
YCbCr	85.52	9.29	14.48
normalized RGB	92.9	2.18	7.1
HS	85.69	13.5	14.31
CbCr	79.18	20.8	20.82
nRG	82.64	17.1	17.36
HS-CbCr	86.14	13.02	13.86
Proposed	94.13	0.046	0.104

FIGURE 4.28. Accuracy, FPR and FNR for HGR dataset [62]

FIGURE 4.29. Overall performance of HGR dataset [62]

method is in green. HSV achieved the closest accuracy to the proposed algorithm with a difference of less than 1%. For FPR, the proposed method is quite low, almost an order of magnitude less than each of the other methods. This indicates the effectiveness of the proposed method for correct classification of non-skin pixels. The false negative rate is depicted with a black line. From the black line, it is clear the proposed algorithm has very low false negative detection as well. The overall performance is shown in Figure 4.29.

4.2. Summary

The proposed method has been compared with existing skin detection methods. To ensure comprehensive comparisons, Compaq, SFA and HGR datasets and the relevant research papers in which performance results are published are used. The several researchers have used various performance metrics to establish efficacy of the methods each have investigated.

The performance metrics employed overlap between authors but are not exactly the same subsets. Fortunately, the metrics can often be transformed from one to another. All of those data has been adopted from published research. Where necessary mappings among metrics were applied. In the end, the proposed method outperforms each of the identified alternatives. In terms of simplicity, the method is straightforward and takes reasonable execution time. These methods can be adapted to detect skin on videos as well as images. Lastly, experimental results have shown a promising skin detection method that performs sufficiently well to be applied in real-world situations.

CHAPTER 5

CONCLUSION

In this research, a clustering-based skin detection method is proposed and validated. The experimental results are compared to various previously published research papers. Three datasets, Compaq, SFA, and HGR, was used to evaluate skin detection. A broad array of performance metrics were computed and compared.

The first chapter of this dissertation contains a detailed introduction to skin detection and motivation for applying it. It focuses on the challenges of skin detection in real world scenarios. The effect of various factors – lighting, ethnic origin, background – on skin detection is described. The motivation for this work is also described. Among the challenges, lighting conditions is the most widely researched and complex. Without the effect of lighting conditions, skin detection can be achieved at an excellent accuracy. Because of the possible applications of a number of human-computer interactions, skin detection still holds much interest among image and video researchers.

The second chapter reviewed different color spaces that are used to represent images. There are many color spaces that have been used to represent and distinguish skin from non-skin pixels. Because each color space has unique characteristics, skin-pixel representation can exhibit different signatures. Thus based on each color space many skin detection methods have been proposed. They can be divided into pixel-based and region-based methods. A pixel-based method considers local information for a pixel and assigns a skin or non-skin label. In that context, The local information can be interpreted using information – statistical, machine learning, and threshold-based – that is extracted via pre-processing.

Statistical-based methods are mostly based on probability theory and may be histogram-based. These methods are trained and then create a model that is then used for skin detection. One broad class, parametric methods, use skin-pixel distributions in the image. Most machine learning based approaches use artificial neural networks or clustering. The former requires much training data to accurately classify skin pixels within an image. Nonetheless,

in the end, threshold-based methods are popular for different color spaces. The threshold is based on placing tight bounds on one or more channels of the color space. These techniques perform well but they are not robust due to limitless possible variations of conditions (including illumination).

Until now, there does not exist a method that can efficiently detect skin while overcoming all challenges. It is quite obvious that a robust and efficient method is on high demand. The proposed method is robust and experimentally validated and it has been validated via the most used, publicly available datasets.

The third chapter describes the proposed method. It is an unsupervised method using clustering. Unsupervised approaches are not as broadly used for skin detection. Also, supervised methods require quality training data to build a model that can perform well on testing data. Yet, availability of all types of skin data for training is unrealistic. On the other hand, unsupervised methods are free from a training phase and work by automatically selecting the key skin features. We speak here of clustering. The main issue for unsupervised clustering is that the number of clusters must be known in advance.

Obviously, for skin detection, a fixed number of clusters cannot be chosen then applied for all images. Thus, the number of clusters must be assigned dynamically. This is among the main contributions of this research. Based on the number of objects present in the image, the number of clusters are assigned dynamically for each image. Experimental results validate the technique. These experiments were performed not only for the three datasets but also for other publicly available datasets for fair comparisons. Moreover, this technique can also be applied to other two-class segmentation problems such as cloud vs. background or vegetation vs. background. It opens windows for a researcher interested in the same research domain as well.

Nonetheless, once a set of skin clusters are identified, there exist challenging problems in evaluating them. Another finding of this research is to obtain a range of hue values for skin. Experimental evaluation for people of various ethnic groups has proven the claim that, skin pixels fall into a fixed range of hue values across groups. This research technique, but

not the specific result, opens a door for new research areas to which it could be applied. However, due to unsupervised nature of clustering methods, some skin pixels are missed for some real-world conditions. Not all of them will fall into a “skin cluster.” This limitation can be overcome by using a region-growing procedure which is also developed (but, admittedly, worthy of further refinement). In addition to that, combining more than one method might result in a yet more efficient skin detection classifier.

Applications of skin detection range broadly from gesture recognition, face detection, pose analysis to the several sorts of disease detection in medicine. A good skin detector can lead to applications that assist people with special needs in using computer-based applications such as gesture-based alternatives to human-computer interfaces. This might include recognition of head/hand gestures, such as those used in American sign language. Athletes can be benefited through examination of their performance of various movements and suggestions for improvements where needed.

In summary, the proposed algorithm shows promising results but there is room for improvement. Undoubtedly the convolution neural network (CNN) approach has superior performance for many classification problems. As mentioned earlier CNNs must be trained with much data and all possible data variations. Given the proposed algorithm can detect skin efficiently and effectively, the results can be used to train a CNN. For a CNN, each pixel will be labeled as skin and non-skin to serve as ground truth for training. Previous CNN-based experience assures that accurate training images result in accurate classification results. For this problem domain, that should be achievable. On other hand, an alternative plan might also be considered. Ensemble classification is a possibility. In ensemble classification, several classifiers examine an instance and each votes on a prediction. Often each classifier’s vote is given a weight based on importance of the subset of selected features examines. Ensemble classifiers can definitely be applied to the goal of skin detection and are attractive because no single classifier is sufficient to overcome all the problem’s challenges.

REFERENCES

- [1] HANI K Al-Mohair, JM Saleh, and SA Saundi, *Impact of color space on human skin color detection using an intelligent system*, 1st WSEAS International Conference on Image Processing and Pattern Recognition (IPPR'13), 2013.
- [2] Hani K Al-Mohair, Junita Mohamad Saleh, and Shahrel Azmin Suandi, *Hybrid human skin detection using neural network and k-means clustering technique*, Applied Soft Computing 33 (2015), 337–347.
- [3] Alberto Albiol, David Monzo, Antoine Martin, Jorge Sastre, and Antonio Albiol, *Face recognition using hog–ebgm*, Pattern Recognition Letters 29 (2008), no. 10, 1537–1543.
- [4] David Arthur and Sergei Vassilvitskii, *k-means++: The advantages of careful seeding*, Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
- [5] Jorge Alberto Marcial Basilio, Gualberto Aguilar Torres, Gabriel Sánchez Pérez, L Karina Toscano Medina, and Héctor M Pérez Meana, *Explicit image detection using ycbcr space color model as skin detection*, Applications of Mathematics and Computer Engineering (2011), 123–128.
- [6] Michael JA Berry and Gordon S Linoff, *Data mining techniques: for marketing, sales, and customer relationship management*, John Wiley & Sons, 2004.
- [7] N Brancati, M Frucci, G De Pietro, and L Gallo, *Dynamic clustering for skin detection in ycbcr colour space*, (2016).
- [8] Nadia Brancati, Giuseppe De Pietro, Maria Frucci, and Luigi Gallo, *Human skin detection through correlation rules between the ycb and ycr subspaces based on dynamic color clustering*, Computer Vision and Image Understanding 155 (2017), 33–42.
- [9] Jason Brand and John S Mason, *A comparative assessment of three approaches to pixel-level human skin-detection*, Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol. 1, IEEE, 2000, pp. 1056–1059.
- [10] David A Brown, Ian Craw, and Julian Lewthwaite, *A som based approach to skin detection with application in real time systems.*, BMVC, vol. 1, Citeseer, 2001, pp. 491–500.

- [11] Tiberio S Caetano, Silvia D Olabarriaga, and Dante Augusto Couto Barone, *Performance evaluation of single and multiple-gaussian models for skin color modeling*, Proceedings. XV Brazilian Symposium on Computer Graphics and Image Processing, IEEE, 2002, pp. 275–282.
- [12] Tiberio S Caetano, Silvia D Olabarriaga, and Dante AC Barone, *Do mixture models in chromaticity space improve skin detection?*, Pattern Recognition 36 (2003), no. 12, 3019–3021.
- [13] Joao Paulo Brognoni Casati, Diego Rafael Moraes, and Evandro Luis Linhari Rodrigues, *Sfa: A human skin image database based on feret and ar facial images*, IX workshop de Visao Computational, Rio de Janeiro, 2013.
- [14] Douglas Chai and King N Ngan, *Face segmentation using skin-color map in videophone applications*, IEEE Transactions on circuits and systems for video technology 9 (1999), no. 4, 551–564.
- [15] Douglas Chai, Son Lam Phung, and Abdesselam Bouzerdoun, *Skin color detection for face localization in human-machine communications*, Proceedings of the Sixth International Symposium on Signal Processing and its Applications (Cat. No. 01EX467), vol. 1, IEEE, 2001, pp. 343–346.
- [16] Jose M Chaves-González, Miguel A Vega-Rodríguez, Juan A Gómez-Pulido, and Juan M Sánchez-Pérez, *Detecting skin in face recognition systems: A colour spaces study*, Digital Signal Processing 20 (2010), no. 3, 806–823.
- [17] Byeongcheol Choi, Seungwan Han, Byungho Chung, and Jaecheol Ryou, *Human body parts candidate segmentation using laws texture energy measures with skin color*, 13th International Conference on Advanced Communication Technology (ICACT2011), IEEE, 2011, pp. 556–560.
- [18] Ingo Dahm, Sebastian Deutsch, Matthias Hebbel, and André Osterhues, *Robust color classification for robot soccer*, 7th International Workshop on RoboCup, 2003.
- [19] Djamila Dahmani and Slimane Larabi, *User-independent system for sign language finger spelling recognition*, Journal of Visual Communication and Image Representation 25

- (2014), no. 5, 1240–1250.
- [20] Thierry Denoeux, *A neural network classifier based on dempster-shafer theory*, IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 30 (2000), no. 2, 131–150.
 - [21] CN Deshmukh and Ms SP Wankhade, *An efficient algorithm for face detection using color segmentation and energy thresholding*, International Journal 1 (2013).
 - [22] Richard O Duda, Peter E Hart, and David G Stork, *Pattern classification*, John Wiley & Sons, 2012.
 - [23] Joseph C Dunn, *Well-separated clusters and optimal fuzzy partitions*, Journal of cybernetics 4 (1974), no. 1, 95–104.
 - [24] CE Erdem, Sezer Ulukaya, Ali Karaali, and A Tanju Erdem, *Combining haar feature and skin color based classifiers for face detection*, 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2011, pp. 1497–1500.
 - [25] Rodrigo Augusto Dias Faria and Roberto Hirata Jr, *Combined correlation rules to detect skin based on dynamic color clustering.*, VISIGRAPP (5: VISAPP), 2018, pp. 309–316.
 - [26] Christophe Garcia and George Tziritas, *Face detection using quantized skin color regions merging and wavelet packet analysis*, IEEE Transactions on multimedia 1 (1999), no. 3, 264–277.
 - [27] Francesca Gasparini, Silvia Corchs, and Raimondo Schettini, *Recall or precision-oriented strategies for binary classification of skin pixels*, Journal of electronic imaging 17 (2008), no. 2, 023017.
 - [28] Giovani Gomez, *On selecting colour components for skin detection*, Object recognition supported by user interaction for service robots, vol. 2, IEEE, 2002, pp. 961–964.
 - [29] Rafael C Gonzalez, Richard E Woods, et al., *Digital image processing [m]*, Publishing house of electronics industry 141 (2002), no. 7.
 - [30] Junwei Han, GM Award, Alistair Sutherland, and Hai Wu, *Automatic skin segmentation for gesture recognition combining region and support vector machine active learning*, 7th International Conference on Automatic Face and Gesture Recognition (FGR06), IEEE,

- 2006, pp. 237–242.
- [31] Eun-Jung Holden, Gareth Lee, and Robyn Owens, *Australian sign language recognition*, Machine Vision and Applications 16 (2005), no. 5, 312.
 - [32] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, and Anil K Jain, *Face detection in color images*, IEEE transactions on pattern analysis and machine intelligence 24 (2002), no. 5, 696–706.
 - [33] Quan Huynh-Thu, Mitsuhiro Meguro, and Masahide Kaneko, *Skin-color extraction in images with complex background and varying illumination*, Sixth IEEE Workshop on Applications of Computer Vision, 2002.(WACV 2002). Proceedings., IEEE, 2002, pp. 280–285.
 - [34] Abm Rezbaul Islam, Ali Alammari, and Bill Buckles, *Skin detection in image and video founded in clustering and region growing*, Mobile Multimedia/Image Processing, Security, and Applications 2019, vol. 10993, International Society for Optics and Photonics, 2019, p. 109930V.
 - [35] Alejandro Jaimes and Nicu Sebe, *Multimodal human–computer interaction: A survey*, Computer vision and image understanding 108 (2007), no. 1-2, 116–134.
 - [36] Zhichao Ji and Huabiao Qin, *A vehicle surveillance system for face detection*, 2007 IEEE International Conference on Vehicular Electronics and Safety, IEEE, 2007, pp. 1–4.
 - [37] Michael J Jones and James M Rehg, *Statistical color models with application to skin detection*, International Journal of Computer Vision 46 (2002), no. 1, 81–96.
 - [38] Praveen Kakumanu, Sokratis Makrogiannis, and Nikolaos Bourbakis, *A survey of skin-color modeling and detection methods*, Pattern recognition 40 (2007), no. 3, 1106–1122.
 - [39] Hyun Kang, Chang Woo Lee, and Keechul Jung, *Recognition-based gesture spotting in video games*, Pattern Recognition Letters 25 (2004), no. 15, 1701–1714.
 - [40] SN Karishma and V Lathasree, *Fusion of skin color detection and background subtraction for hand gesture segmentation*, International Journal of Engineering Research and Technology 3 (2014), no. 2, 1835–1839.
 - [41] S Ketenci and B Gencturk, *Performance analysis in common color spaces of 2d gaussian*

- color model for skin segmentation. ineurocon, 2013 ieee (pp. 1653-1657)*, 2013.
- [42] Rehanullah Khan, Allan Hanbury, Julian Stöttinger, and Abdul Bais, *Color based skin classification*, Pattern Recognition Letters 33 (2012), no. 2, 157–163.
 - [43] Sang-Hoon Kim, Nam-Kyu Kim, Sang Chul Ahn, and Hyoung-Gon Kim, *Object oriented face detection using range and color information*, Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 1998, pp. 76–81.
 - [44] Jure Kovac, Peter Peer, and Franc Solina, *Human skin color clustering for face detection*, vol. 2, IEEE, 2003.
 - [45] Jae Y Lee and Suk I Yoo, *An elliptical boundary model for skin color detection*, Proc. of the 2002 International Conference on Imaging Science, Systems, and Technology, 2002.
 - [46] Gang Li, Yinping Xu, and Jiaying Wang, *An improved adaboost face detection algorithm based on optimizing skin color model*, 2010 sixth international conference on natural computation, vol. 4, IEEE, 2010, pp. 2013–2015.
 - [47] Stephen J McKenna, Shaogang Gong, and Yogesh Raja, *Modelling facial colour and identity with gaussian mixtures*, Pattern recognition 31 (1998), no. 12, 1883–1892.
 - [48] Rosali Mohanty and MV Raghunadh, *Skin color segmentation based face detection using multi-color space*, International Journal of Advanced Research in Computer and Communication Engineering 5 (2016), no. 5, 470–475.
 - [49] J Montenegro, W Gomez, and P Sanchez-Orellana, *A comparative study of color spaces in skin-based face segmentation*, 2013 10th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), IEEE, 2013, pp. 313–317.
 - [50] Yasir M Mustafah, Abbas Bigdeli, Amelia W Azman, and Brian C Lovell, *Face detection system design for real time high resolution smart camera*, 2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC), IEEE, 2009, pp. 1–6.
 - [51] Nadia Nacer, Bouraoui Mahmoud, and Mohamed Hedi Bedoui, *Fpga architecture for facial-features and components extraction*, International Journal of Computer Science (2013).
 - [52] Sinan Naji, Hamid A. Jalab, and Sameem A. Kareem, *A survey on skin detection in*

- colored images*, Artificial Intelligence Review 52 (2019), no. 2, 1041–1087.
- [53] Jakub Nalepa, Tomasz Grzejszczak, and Michal Kawulok, *Wrist localization in color images for hand gesture recognition*, Man-Machine Interactions 3, Springer, 2014, pp. 79–86.
 - [54] Jakub Nalepa and Michal Kawulok, *Fast and accurate hand shape classification*, International conference: beyond databases, architectures and structures, Springer, 2014, pp. 364–373.
 - [55] Nuria Oliver, Alex P Pentland, and Francois Berard, *Lafter: Lips and face real time tracker*, Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 1997, pp. 123–129.
 - [56] Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai, *Skin segmentation using color pixel classification: analysis and comparison*, IEEE Transactions on Pattern Analysis & Machine Intelligence (2005), no. 1, 148–154.
 - [57] Christian Platzter, Martin Stuetz, and Martina Lindorfer, *Skin sheriff: a machine learning solution for detecting explicit images*, Proceedings of the 2nd international workshop on Security and forensics in communication systems, ACM, 2014, pp. 45–56.
 - [58] Rudra PK Poudel, Hammadi Nait-Charif, Jian J Zhang, and David Liu, *Region-based skin color detection.*, VISAPP (1) VISAPP 2012-Proceedings of the International Conference on Computer Vision Theory and Applications 1: 301-306 2012, vol. 1, VISAPP, 2012, pp. 301–306.
 - [59] Varsha Powar, Amruta Kulkarni, Renuka Lokare, and Aishwarya Lonkar, *Skin detection for forensic investigation*, 2013 International Conference on Computer Communication and Informatics, IEEE, 2013, pp. 1–4.
 - [60] Charles Poynton, *Poynton's color faq*, Electronic Preprint (1995).
 - [61] Yong qiang Chen, Changdong Wu, and Bo Yu, *The skin color segmentation algorithm based on the $cr^* a^* b$ color space*, International Journal of Control and Automation 9 (2016), no. 5, 255–262.
 - [62] Romi Fadillah Rahmat, Tengku Chairunnisa, Dani Gunawan, and Opim Salim Sitom-

- pul, *Skin color segmentation using multi-color space threshold*, 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), IEEE, 2016, pp. 391–396.
- [63] Siddharth S Rautaray and Anupam Agrawal, *Vision based hand gesture recognition for human computer interaction: a survey*, Artificial intelligence review 43 (2015), no. 1, 1–54.
- [64] John C Russ and J Christian Russ, *Introduction to image processing and analysis*, CRC press, 2007.
- [65] Alaa Sagheer and Saleh Aly, *An effective face detection algorithm based on skin color information*, 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, IEEE, 2012, pp. 90–96.
- [66] David Saxe and Richard Foulds, *Toward robust skin identification in video images*, Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, IEEE, 1996, pp. 379–384.
- [67] Richard P Schumeyer and Kenneth E Barner, *Color-based classifier for region identification in video*, Visual Communications and Image Processing’98, vol. 3309, International Society for Optics and Photonics, 1998, pp. 189–201.
- [68] Karl Schwerdt and James L Crowley, *Robust face tracking using color*, Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), IEEE, 2000, pp. 90–95.
- [69] Nicu Sebe, Ira Cohen, Thomas S Huang, and Theo Gevers, *Skin detection: A bayesian network approach*, Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., vol. 2, IEEE, 2004, pp. 903–906.
- [70] Haidar R Shakir¹, Loay E George, and Ghadah K Tuama, *Single face detection based on skin color and edge detection*, (2015).
- [71] Linda Shapiro, *Computer vision and image processing*, Academic Press, 1992.
- [72] Wladyslaw Skarbek, Andreas Koschan, Technischer Bericht, Zur Veroffentlichung, et al., *Colour image segmentation-a survey*, (1994).

- [73] Karin Sobottka and Ioannis Pitas, *A novel method for automatic face segmentation, facial feature extraction and tracking*, Signal processing: Image communication 12 (1998), no. 3, 263–281.
- [74] Milan Sonka, Vaclav Hlavac, and Roger Boyle, *Image processing, analysis, and machine vision*, Cengage Learning, 2014.
- [75] Maricor Soriano, Birgitta Martinkauppi, Sami Huovinen, and Mika Laaksonen, *Skin detection in video under changing illumination conditions*, Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol. 1, IEEE, 2000, pp. 839–842.
- [76] Hugo Steinhaus, *Sur la division des corp materiels en parties*, Bull. Acad. Polon. Sci 1 (1956), no. 804, 801.
- [77] Moritz Störring, Tomáš Kočka, Hans J Andersen, and Erik Granum, *Tracking regions of human skin through illumination changes*, Pattern Recognition Letters 24 (2003), no. 11, 1715–1723.
- [78] Ravi Subban and Richa Mishra, *Human skin segmentation in color images using gaussian color model*, Recent Advances in Intelligent Informatics, Springer, 2014, pp. 13–21.
- [79] Wei Ren Tan, Chee Seng Chan, Pratheepan Yogarajah, and Joan Condell, *A fusion approach for efficient human skin detection*, IEEE Transactions on Industrial Informatics 8 (2011), no. 1, 138–147.
- [80] San Tang, *Human face detection method based on skin color model*, Advanced Materials Research, vol. 706, Trans Tech Publ, 2013, pp. 1877–1881.
- [81] Pooya Tavallali, Mehran Yazdi, and Mohammad Reza Khosravi, *Robust cascaded skin detector based on adaboost*, Multimedia Tools and Applications 78 (2019), no. 2, 2599–2620.
- [82] Prahlad Vadakkepat, Peter Lim, Liyanage C De Silva, Liu Jing, and Li Li Ling, *Multi-modal approach to human-face detection and tracking*, IEEE transactions on industrial electronics 55 (2008), no. 3, 1385–1393.
- [83] Vladimir Vezhnevets and Alla Andreeva, *A comparative assessment of pixel-based skin detection methods*, GML Computer Vision Group (2005).

- [84] Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva, *A survey on pixel-based skin color detection techniques*, Proc. Graphicon, vol. 3, Moscow, Russia, 2003, pp. 85–92.
- [85] Dong Wang, Jinchang Ren, Jianmin Jiang, and Stan S Ipson, *Skin detection from different color spaces for model-based face detection*, International Conference on Intelligent Computing, Springer, 2008, pp. 487–494.
- [86] Jian-Gang Wang and Eric Sung, *Frontal-view face detection and facial feature extraction using color and morphological operations*, Pattern recognition letters 20 (1999), no. 10, 1053–1068.
- [87] Yanjiang Wang and Baozong Yuan, *A novel approach for human face detection from color images under complex background*, Pattern Recognition 34 (2001), no. 10, 1983–1992.
- [88] Yanwen Wu and Xueyi Ai, *Face detection in color images using adaboost algorithm based on skin color information*, First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), IEEE, 2008, pp. 339–342.
- [89] Fan Hai Xiang and Shahrel Azmin Suandi, *Fusion of multi color space for human skin regionsegmentation*, International Journal of Information and Electronics Engineering 3 (2013), no. 2, 172.
- [90] Jie Yang, Weier Lu, and Alex Waibel, *Skin-color modeling and adaptation*, Asian Conference on Computer Vision, Springer, 1998, pp. 687–694.
- [91] Ming-Hsuan Yang and Narendra Ahuja, *Gaussian mixture model for human skin color and its applications in image and video databases*, Storage and retrieval for image and video databases VII, vol. 3656, International Society for Optics and Photonics, 1998, pp. 458–467.
- [92] Ming-Hsuan Yang, David J Kriegman, and Narendra Ahuja, *Detecting faces in images: A survey*, IEEE Transactions on pattern analysis and machine intelligence 24 (2002), no. 1, 34–58.
- [93] Nur Baiti Zahir, Rosdiyana Samad, and Mahfuzah Mustafa, *Initial experimental results of real-time variant pose face detection and tracking system*, 2013 IEEE International

- Conference on Signal and Image Processing Applications, IEEE, 2013, pp. 264–268.
- [94] AA Zaidan, Nurul Nadia Ahmad, H Abdul Karim, Moussa Larbani, BB Zaidan, and Aduwati Sali, *On the multi-agent learning neural and bayesian methods in skin detector and pornography classifier: An automated anti-pornography system*, Neurocomputing 131 (2014), 397–418.
- [95] Benjamin D Zarit, Boaz J Super, and Francis KH Quek, *Comparison of five color models in skin pixel classification*, Proceedings International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. In Conjunction with ICCV’99 (Cat. No. PR00378), IEEE, 1999, pp. 58–63.
- [96] Qiang Zhu, Kwang-Ting Cheng, Ching-Tung Wu, and Yi-Leh Wu, *Adaptive learning of an accurate skin-color model*, Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings., IEEE, 2004, pp. 37–42.
- [97] Haiqiang Zuo, Heng Fan, Erik Blasch, and Haibin Ling, *Combining convolutional and recurrent neural networks for human skin detection*, IEEE Signal Processing Letters 24 (2017), no. 3, 289–293.