A Novel Framework for Segmentation of Skin Regions using Chromatic and Textural Information

A

Thesis Submitted

in Fulfilment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Biplab Ketan Chakraborty

Department of Electronics and Electrical Engineering

Indian Institute of Technology Guwahati

Guwahati, India.

March, 2018

In memory of

My Grandmother

Dedicated to

My Loving Wife,

My Beloved Grandfather,

and My Parents

Certificate

This is to certify that the thesis entitled "**A Novel Framework for Segmentation of Skin Regions using Chromatic and Textural Information**", submitted by **Biplab Ketan Chakraborty** (126102023), a research scholar in the *Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati*, for the award of the degree of **Doctor of Philosophy**, has been carried out by her under my supervision and guidance. The thesis has fulfilled all requirements as per the regulations of the institute and in my opinion has reached the standard needed for submission. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dated:

Guwahati.

Dr. M.K. Bhuyan
Associate Professor,
Dept. of Electronics and Electrical Engineering,
Indian Institute of Technology Guwahati,
India - 781039.

Acknowledgements

I would like to thank all those people who made this dissertation possible.

First of all, I would like to express my profound respect and gratitude to my supervisor, Dr. M.K. Bhuyan, who has been the guiding force behind this work. I am greatly indebted for his guidance, constant encouragement, and valuable comments on my work. I am fortunate enough to have such an advisor who gave me the freedom to think independently and explore new ideas. More importantly, I would like to thank for the patience he has shown in carefully reading and commenting on the manuscripts, and countless revisions of this dissertation. His commitments and dedication to research have been and will continue to be a constant source of inspiration for me. I have no doubts that finishing my degree in a proper and timely manner would have been impossible without his help. I am highly privileged for getting an opportunity to work with such a wonderful person.

I would also like to thank my doctoral committee members Prof. P. K. Bora, Dr. S. Sundaram and Dr. S. Bhattacharya for their invaluable suggestions, encouragements, and moral supports that helped me to improve my research work. I am also thankful to the Head, other faculty members and non-teaching staff of Department of Electronics & Electrical Engineering for their kind help extended during my academic studies.

My special thanks to Mr. Sanjib Das for providing necessary resources useful for the research work. My special thanks to my friends-cum-moral adviser Mr. Sumantra Chaudhuri and Dr. Somak Bhattacharya, for their guidance and insightful comments during the entire journey of my PhD life.

On a personal note, I would like to thank my friend Dr. Sunil Kumar for his constant support and for being with me in every aspect of my life. My special thanks go to Mr. Santhosh, Mr. Parveen, Mr. Gaurav, Mr. Debajit, Mr. Pradipta, Mr. Tilendra, Mr. Aniruddha, Mr. Shakhanil, and Mr. Soumayan, for their motivation and support. I had a great time with many of my friends at IIT Guwahati, including (but not limited to) Mr. Mandar Mitra, Mr. Himangshu Bhaumik, and Mr. Arijit Roy. I would like to thank them for their support and encouragement.

I am grateful to my grandparents, parents, my wife, and my in-laws whose love encouragement, and support made this research work possible. I am thankful to IIT Guwahati for providing the research scholarship to undertake my PhD research. A kind thanks to all the doctors, nurses, and staff members of Institute Hospital who have taken care of my health timely. Finally, I would like to thank the Almighty God for bestowing me this opportunity and showering his blessings on me to come out successful against all odds.

Biplab Ketan Chakraborty

Abstract

Skin detection is an important step in various image processing and vision-based Human-Computer Interaction (HCI) applications. It is the process of finding skin-coloured pixels and regions in an image or a video. Segmenting skin from real-world images is a difficult task even though human skin is known to possess a unique color range. The major challenges of skin detection in images are – presence of skin-like colours in background and changes in chromatic appearance of skin regions due to non-uniform illumination. In addition to these problems, detection of skin regions in videos is more challenging in presence of time-varying illumination conditions and dynamic backgrounds. Motivated by these facts, we have proposed a set of skin detection algorithms for different environmental conditions using chromatic and textural properties of skin regions.

Standard skin probability map (SPM) can not perfectly discriminate skin and non-skin regions in the above mentioned conditions. To overcome limitations of SPM, a new probability map termed as discriminative space map (DSM) is proposed by extracting most discriminative features between skin and non-skin regions. A novel adaptive discriminative analysis (ADA) is proposed to extract most discriminant features between skin and non-skin regions from an image itself in an unsupervised manner. Subsequently, a dynamic region growing (DRG) method is employed to allow skin regions to grow dynamically. The DRG controls false detection by restricting the region growing process. Experimental results show that the proposed method can efficiently segment out skin pixels in uncontrolled environments.

Our earlier proposed method assumes that the biggest skin-coloured region in an image belongs to actual skin region. Also, our earlier proposed method cannot perfectly handle non-uniform illumination as improper illumination makes skin regions to appear darker than its actual tone. To address these issues, a novel skin detection method is proposed by utilising an image pixel distribution model (IDM), which is derived using a Gaussian Mixture Model (GMM) in a given colour space. In this method, a local skin distribution model (LSDM) and a local background distribution model (LBDM) are derived by exploiting the similarity between the IDM and a reference skin pixel distribution model. The reference skin model is derived from a set of facial skin pixels, and it is termed as facial skin distribution model (FSDM). A local skin probability map (LSPM) can be derived using the LSDM and the LBDM. Finally, a fusion-based skin probability map (FSPM) is obtained by using both the LSPM and an SPM derived from globally obtained skin and non-skin training samples. Subsequently, the proposed DRG algorithm is applied on the FSPM for further reduction in detection errors. Experimental results show that the proposed FSPM can better discriminate skin regions from non-skin regions as compared to state-of-the-art methods for skin detection in images.

The proposed FSPM-based skin segmentation method is applicable for a video when the background is static and the ambient illumination does not change for all the frames of the video. However, the chromatic appearance of skin regions may change locally due to local shading effects on account of the motion of body parts. To address this specific issue, a dynamic adaptation scheme is proposed to detect skin regions which are affected by local colour deformations. The proposed method has two modules – a static module for detection of static skin regions; and a dynamic module for detection of moving skin regions. The static module consists of an FSDM and a video specific background model. The video specific background model termed as fusion-based background distribution model (FBDM) is obtained using an LBDM and a global background distribution model (GBDM). The LBDM for this method is obtained by considering the similarities of a frame pixel distribution model (equivalent to IDM) with the FSDM and the

GBDM. Subsequently, the FBDM is derived by using the LBDM and the GBDM. In the dynamic module of the proposed method, a moving skin distribution model (MSDM) is derived from a set of moving skin samples. Initially, the moving skin regions are detected using a modified double frame-difference method. An initial model for these moving skin regions is obtained by using a GMM. However, some background regions can also be falsely detected as skin regions during the moving object detection process. The final MSDM is obtained by performing a filtering procedure based on similarities of the initial moving skin model with the FSDM and the FBDM. Finally, the static and the dynamic modules are fused by following a maximisation rule. Experimental analysis shows that the proposed skin detection method can detect skin region more accurately than the state-of-the-art methods when there exist local chromatic variations of skin appearance.

Contents

List of Figures

List of Tables

List of Acronyms

ADA	Adaptive Discriminative Analysis
ASSC	Adaptive Seed-based Skin Classification
BIC	Bayesian Information Criterion
DRG	Dynamic Region Growing
DSM	Discriminative Space Map
DSMSC	Dynamic Skin Model-based Skin Classification
DSPF	Discriminative Skin-Presence Feature
DTSC	Dynamic Threshold–based Skin Classification
FNT	Flexible Neural Tree
FBDM	Fusion-based Background Distribution Model
FPSS	Fast Propagation-based Skin Segmentation
FSDM	Facial Skin Distribution Model
FSPM	Fusion-based Skin Probability Map
GBDM	Global Background Distribution Model
GMM	Gaussian Mixture Model
GSDM	Global Skin Distribution Model
GSPM	Global Skin Probability Map
IDM	Image-pixel Distribution Model
LBDM	Local Background Distribution Model
LDA	Linear Discriminative Analysis
LSDM	Local Skin Distribution Model
LSPM	Local Skin Probability Map

MMSC	Multi Manifold-based Skin Classification
MSDM	Moving Skin-pixel Distribution Model
PCA	Principal Component Analysis
PFCM	Possibilistic Fuzzy C-Means
SASC	Stacked Autoencoders-based skin classification
SASS	Self-adaptive Seed Selection
SHSC	Smooth Histogram-based Skin Classification
RFSC	Random Forest-based Skin Classification

List of Symbols

E_s	Skin-chroma entropy
E	Chromatic entropy
$C(x)$	Cost at location x
$D(\bullet)$	Distance operator
d	Distance value
$G(\bullet)$	Gaussian function
K	Number of Gaussian components in a Gaussian Mixture Model
θ	Threshold for skin probability map
δ	Detection error rate
P	Probability or SPM value
$\mathbf{f}$	Feature Vector
$\mathbf{X}$	Colour vector

1

Introduction

Skin detection plays a very significant role in various image processing and Human-Computer interaction (HCI) applications. In reality, human skin colour has a narrow range of tone, and that is why they cluster in a narrow region in different colour spaces. This unique colour characteristic of human skin makes colour-based skin segmentation very useful. This thesis contributes to the research and development of skin detection in images and videos from three aspects: a) fusion of local skin model derived using facial pixels with another skin model derived using global skin samples, b) extraction of image specific discriminative feature to derive a discriminative feature space for the skin and the non-skin pixels, and c) extension of the first method for videos and time-varying local colour deformations. Experimental results on publicly available databases show that the effectiveness of the proposed approaches for the skin detection in different environments. This chapter gives an overview of the skin detection problem, its applications, and the major challenges in this research. Finally, the organization of the thesis is presented at the end.

1.1 Skin Segmentation

Skin colour and textures are vital signs that individuals utilize intentionally or unknowingly to deduce assortment of culture-related viewpoints about each other. They can be an indication of race, well-being, age, riches, magnificence, and so on [24]. However, these interpretations are subjective and vary over different cultures and ethnicities. Skin segmentation implies automatic detection of skin-coloured like regions in images and videos. Researchers showed significant attention towards skin colour information owing to its computational efficiency, robustness against rotations, scaling and partial occlusions. Figure 1.1 shows a process to detect skin

(a) (b) (c)

Figure 1.1: Skin segmentation: a) original image, b) skin mask and c) segmented image.

coloured regions. Accurate detection of skin regions is a very challenging research problem due to various factors, such as[1]:

- *Illumination:* Skin colour appearance may change significantly under varying illumination conditions (*e.g.* indoor, outdoor, highlights, shadows, non-white lights). This may cause serious degradation in overall detection performance.

- *Skin-like background:* The presence of skin-like colours in the background degrades the detection performance. The non-skin regions may be falsely classified as skin regions due their colour resemblance to actual skin tone.

- *Camera characteristics:* The skin colour distribution of a person may look different in images captured by different cameras even in a standard illumination condition. This is due to the fact that sensor characteristic changes with illumination.

[1]This work has been published in *IET Computer Vision, 2017.* (Refer item 2 in Page 117 for details)

- *Ethnicity:* Human skin colour varies significantly among persons belonging to different ethnic groups or different regions. For example, an Asian person has lighter skin tone than a typical African person. On the other hand, Caucasian people have more white off reddish skin tone.

- *Individual characteristics:* Skin tone of a person also varies with age. Also, the skin tone may changes due to excessive or extensive sun-exposure of body parts.

Skin detection and segmentation is an important research problem as it has many diverse applications, such as:

i) Detection and tracking of various body parts, such as hands and face for gesture recognition and human-computer interaction [19, 25, 26].

ii) Objectionable content detection and filtering of web-based multimedia [27, 28].

iii) Feature extraction for content-based image retrieval [29].

iv) Region of interest-based image coding [30–33], and many more.

According to Kawulok [8], there is a difference in the process of "skin detection" and "skin segmentation". Skin detection is a process where a pixel is treated as an independent entity and it is classified as skin or non-skin based on its chromatic properties. A skin detection process acts as a preliminary step to a skin segmentation process, where a higher level analysis of different regional properties of an image are used for a more accurate extraction of actual skin regions.

At first, this chapter gives a detailed overview of different components of a skin detector. Section 1.1.1 gives a detailed discussion about different colour spaces for skin detection, their advantages and disadvantages. Section 1.2 reviews different classifiers and their applicability in skin detection. The standard evaluation procedure is discussed in Section 1.3. A detailed overview of skin colour modelling algorithms and their extensions to achieve accurate skin segmentation is presented in Section 1.4. This is followed by research motivation in Section 1.5, thesis objectives in Section 1.6, and finally, thesis organization is given in Section 1.7.

1.1.1 Colour spaces for skin detection

Human skin tone appearance is formed by a combination of red colour of blood and melanin (brown, yellow). This unique combination makes the human skin to have a limited range of hues with moderate saturation [34]. Therefore, human skin colour does not scatter randomly in a given colour space, but cluster in a minimal region of a given colour space. This unique colour characteristic of human skin makes colour-based skin segmentation very useful. However, the compactness of the skin coloured clusters is not the same for all the colour spaces. Multiple colour spaces are investigated in literature for skin detection. The choice of colour spaces affects the shape of the skin cluster. This eventually affects the detection accuracy. The discussion is intended to highlight the following points:

- For a given skin patch, the cluster location in a given color space.
- For a given skin patch, the effect of varying illumination on its cluster location in a given colour space.
- The relation between the clusters, derived from a set of given skin patches belonging to different people from the same race in a given colour space.

The answers to the above queries can be illustrated using Figures 1.2 and 1.3. Figure 1.2 shows 2D density plots for skin pixels for Asian people for different colour spaces. On the other hand, Figure 1.3 shows density plots for skin pixels for different races, such as Asian, African and Caucasian plotted for different colour spaces.

1.1.2 RGB colour space

RGB is the most frequently used colour space for digital images. A colour can be represented as an additive combination of three primary colours: Red (R), Green (G) and Blue (B). The RGB colour space can be visualized as a 3-dimensional coordinate space where, R, G, B are the three mutually orthogonal axes. The major advantage of the RGB space is its availability, as almost all the visual light image sensors produce images in RGB colour. This makes RGB as the most primitive colour space in Computer Vision. However, the colour representation is not segregated into luminance and chrominance components, and all the three colour components R, G, B are highly correlated to each other. For example, changing the luminance component

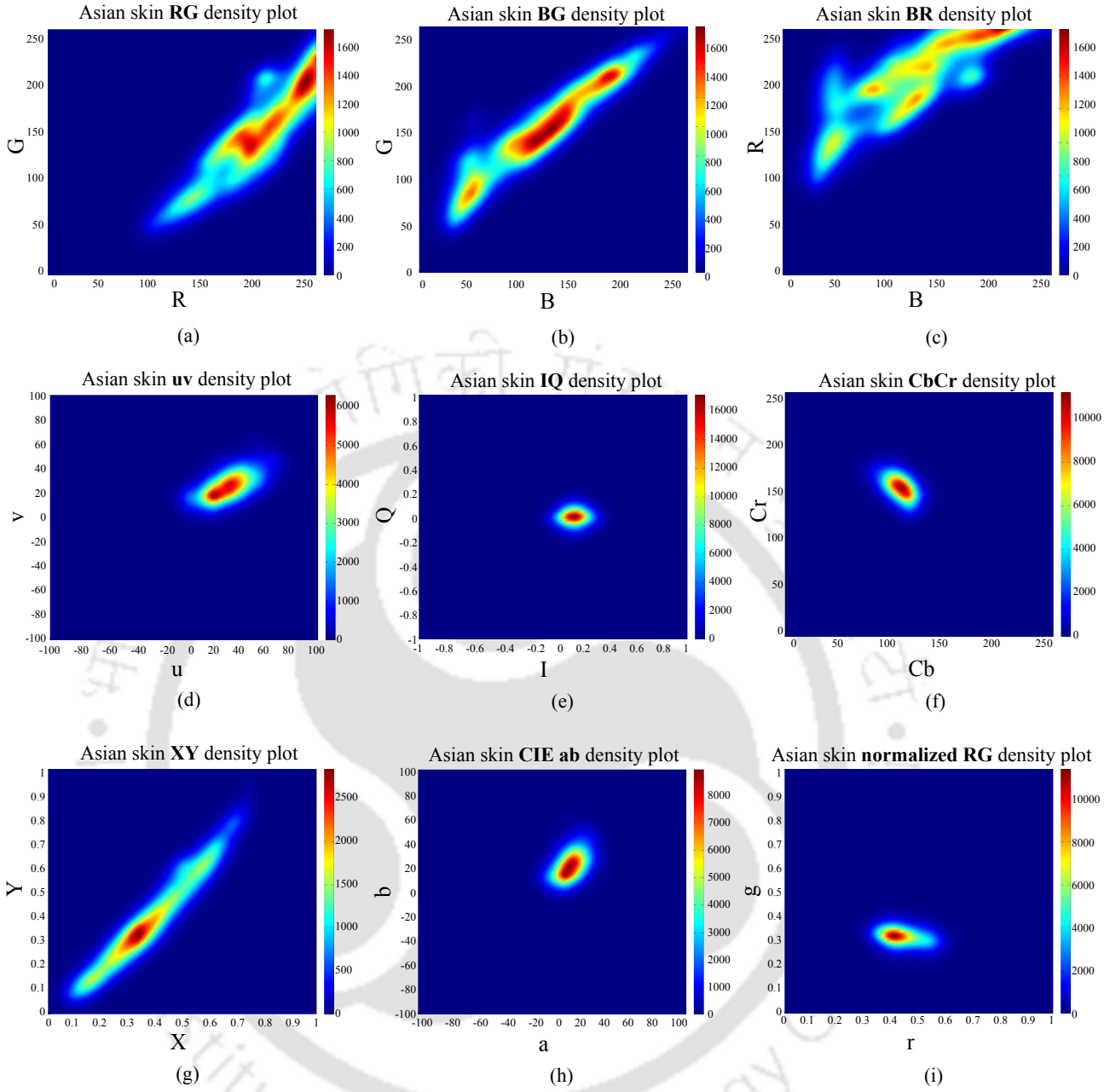


Figure 1.2: Density plots of Asian skin for different color spaces [1]

(average of the three colours) also changes the RGB values of pixels. Therefore, a varying illumination can affect all the RGB elements of a pixel belonging to a skin patch. This eventually results in a change of location of the skin cluster in the RGB space. For example, the plots in the first row of Figure 1.2 are obtained from a set of skin pixels belonging to Asian people under random illumination conditions. The spread of skin-colour cluster in RGB space shows the effects of illumination change on pixel's RGB components. On the other hand, skin colour varies significantly among different races. The primary reason for skin tone variation is due to

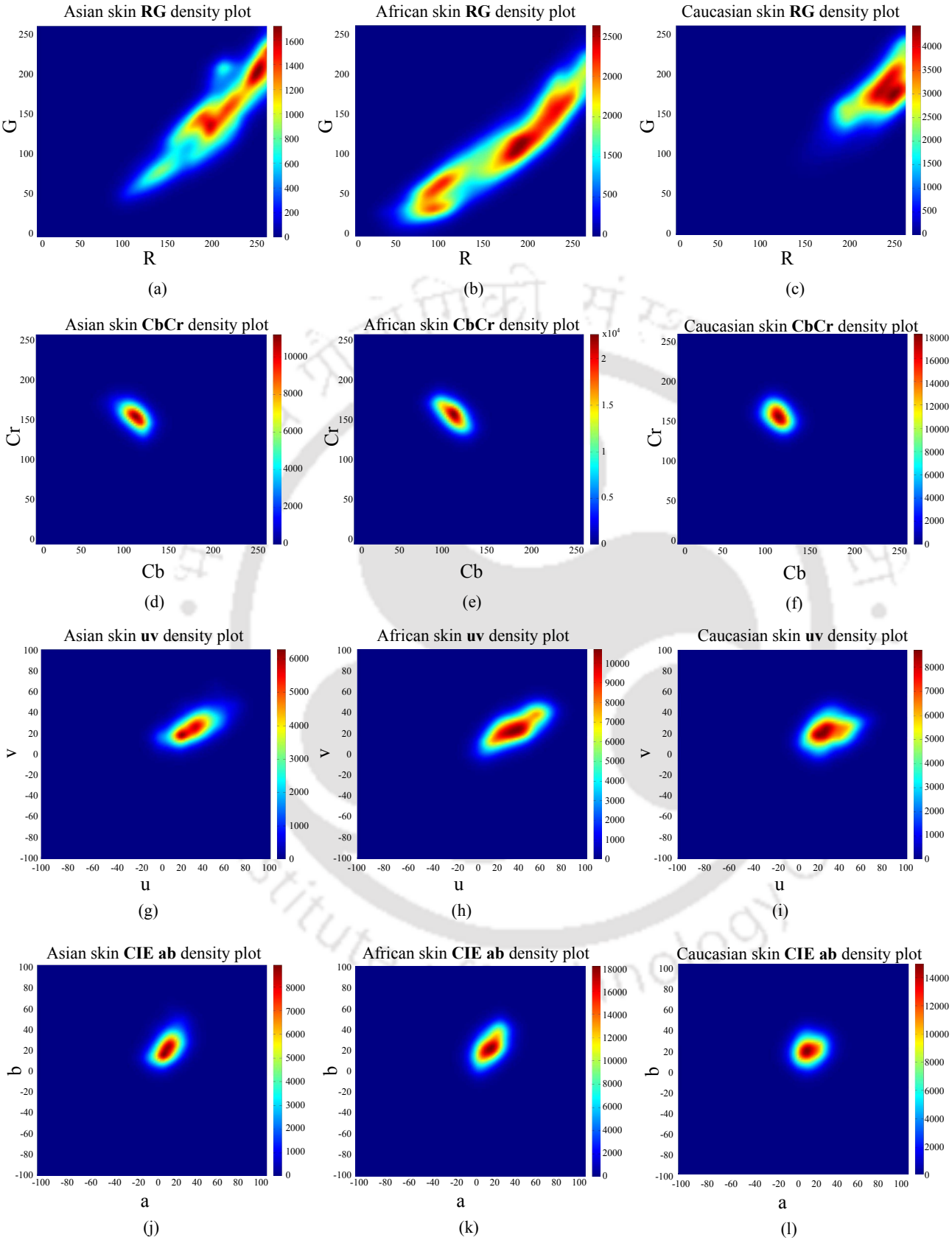


Figure 1.3: Density plots of Asian, African and Caucasian skin for different colour spaces [1]

the variations of the amount of melanin present in the skin of people belonging to different races. For example, African people have more melanin concentration than Caucasian people. Hence, as shown in the first row of Figure 1.3, the skin-colour clusters of different races locate in different locations in a given colour space. Despite these fundamental limitations, the RGB colour space is extensively used in skin detection techniques because of its simplicity. For example, Jones and Rehg [13] used RGB colour space for skin detection, and their experimental results are treated as a benchmark in the domain of skin detection.

1.1.3 TV colour spaces

In television broadcasting technology, orthogonal colour spaces are used, such as YUV, YIQ, and YCbCr colour spaces. Out of these colour spaces, YUV is employed in NTSC TV broadcasting, YUV is used for video encoding and YIQ is used in JPEG image compression and MPEG video compression. All these colour spaces can be obtained using a linear transformation of RGB space.

1.1.3.1 YIQ

YIQ colour model is the NTSC standard for analog video transmission. In YIQ colour model, Y stands for intensity, I is the in-phase component along the orange-cyan axis, and Q is the quadrature component along the magenta-green axis. The Y component is decoupled as the signal has to be made compatible to both monochrome and colour television. The relationship between the YUV and RGB model is given below [35]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1.1}$$

The YIQ model is designed to take advantage of human visual system's greater sensitivity to changes in luminance than to the changes in hue or saturation. Luminance is proportional to the amount of light perceived by the eye. So, the importance of YIQ model is that the luminance component of an image can be processed without affecting its colour content.

1.1.3.2 YCbCr

This is an international standard for studio-quality video. In this, Y is the intensity corresponding to YIQ model while Cb and Cr are so selected that the resulting scheme is efficient for compression (Less spectral redundancy between Cb and Cr, *i.e.*, coefficient are less correlated). This colour model is derived in such a way that it achieves maximum decorrelation. The relationship between the YCbCr and RGB model is listed as follows:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.504 & 0.098 \\ -0.148 & -0.291 & 0.439 \\ 0.439 & -0.368 & -0.071 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1.2}$$

Some of the major reasons for using YCbCr colour space in skin segmentation are given as follows:

- In YCbCr, the luminance component (Y) is independent of chromatic components. This makes the YCbCr colour space suitable for handling the illumination variation problem in skin detection.
- The skin colour cluster shows more compactness in YCbCr space as compared to other spaces [4].
- In YCbCr space, skin and non-skin colours show minimum overlap under different illumination conditions [36].

The YCbCr colour space has been frequently used by many researchers in skin colour detection, such as Chai and Bouzerdoum [37], Phung *et al.* [38], Mahmoud [39], Basilio et al. [40], just to mention a few.

1.1.3.3 YUV

The luminance component (Y) is independent of chromatic components U and V to reduce the effect of lighting variations. The relationship between the YUV and RGB model is given below [35]:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.257 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & 0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1.3}$$

Some of the skin detection methods using YUV colour space are given in [35, 41, 42].

One of the major advantages of using these colour spaces for skin detection is that the majority of multimedia content (image/video) are already encoded using one of them. The ability of these colour spaces in separating the luminance component (Y) from the orthogonal dominance components ($U-V$, $I-Q$, $Cb-Cr$) can help in reducing the effect of varying illumination conditions on skin colour appearance during skin detection. Skin colours typically form a compact cluster with an elliptical shape in 2D chromatic spaces as shown in Figure 1.2-d,e,f. So, neglecting the luminance component makes a skin detector less susceptible to illumination variations. From the shapes of the density plots in 2D chrominance spaces, it is evident that the densities can be easily modelled as a multivariate Gaussian distribution for a single race. It is also observed from Figure 1.3 that the skin colour clusters belonging to different races show significant overlapping with one another. Hence, skin detectors, which are based on these colour spaces can perform well irrespective to the races.

1.1.4 Perceptual colour spaces

In skin detection, another type of colour spaces are used, which are termed as perceptual color space, such as HSI, HSV/HSB, and HSL(HLS). In these colour spaces, colour is represented using three components Hue (H), Saturation (S), and brightness (I, V, or L). The main difference of these colour spaces with TV transmission-based colour spaces, such as YCbCr, YIQ is that the former colour spaces are obtained by using a non-liner transformation from RGB space, whereas TV transmission-based colour spaces use linear transformation. One of the major advantages of using these colour spaces in skin detection is that a skin colour boundary with respect to H and S can be specified intuitively by a user. The brightness components I, V or L are independent of chromatic components H and S. Hence, they can be dropped to reduce the effect of illumination change on skin colour during skin detection. Some of the skin detection methods which use these colour spaces are given in [43–45].

1.1.5 Colorimetric colour spaces

The luminance-chrominance separation can also be achieved by using International Commission on Illumination (Commission Internationale d Eclairage - CIE) defined colour spaces, such as CIE-XYZ, CIE-xy, CIE-Lab. The CIE-XYZ colour space was specified in 1920 as one of

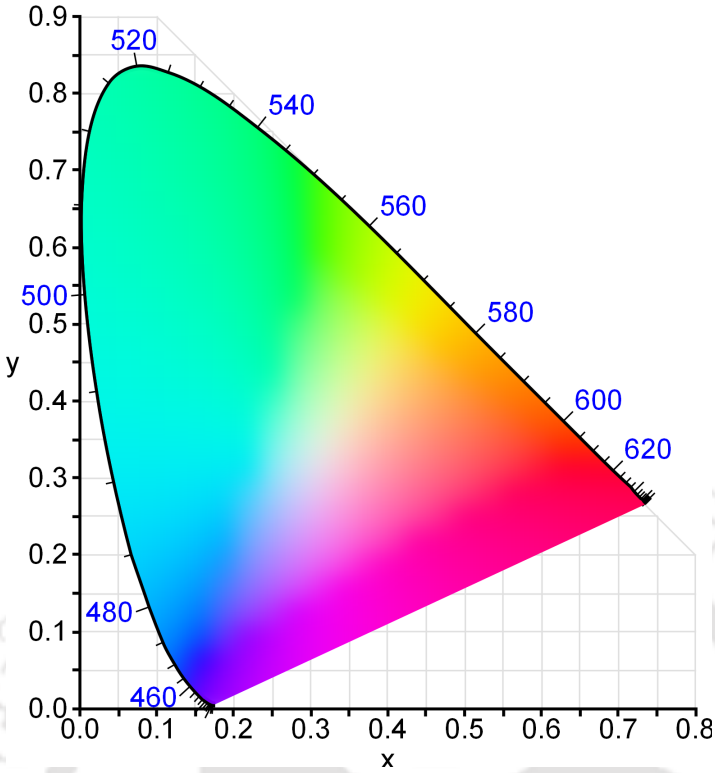


Figure 1.4: Chromaticity diagram for CIE-xy colour space. The boundary shows the spectral (or monochromatic) locus with wavelength in nanometres (picture courtesy : WIKIPEDIA)

the primitive mathematically defined colour space. The colour model is based on human visual perception, and it is used as the foundation for other colorimetric spaces. It can be derived from the RGB colour space following a linear coordinate transformation. In CIE-XYZ, the Y component represents the luminance whereas X and Z represent the chromaticity components. The values of X and Y can be derived by a central projection into the plane $X + Y + Z = 1$ followed by a projection into the XY plane. This results in a horse-shoe-shaped chromaticity diagram in CIE-xy plane as shown in Figure 1.4. One of the limitations of the CIE-XYZ and CIE-xy colour spaces is that the differences in colour are not perceived uniformly over the colour space. The CIE-Lab colour model separates the luminance and chrominance components of a colour as L and a-b. The density plots in chromatic space a-b for skin colours of different races are shown in Figure 1.3. Only a few of the skin detection methods [46–49] used CIE colour spaces on account of computational complexity in the transformation from RGB to CIE colour spaces.

1.2 Classifiers for Skin Detection

Many researchers have used different classifiers in classifying skin and non-skin pixels, such as Bayesian classifier, Support Vector Machine (SVM), Artificial Neural Networks (ANN), decision trees like Random Forest, k-Nearest Neighbour (kNN) etc. A brief discussion about these classifiers is given below.

1.2.1 Histogram model with Naïve Bayes classifier

One of the most frequent used classifier in skin detection is a Naïve Bayes Classifier or simply Bayes Classifier. The Bayes classifier classifies a pixel as skin or non-skin if the a posteriori probability of a given pixel,

$$P\left(Skin|\mathbf{X}\right) \geq \theta \tag{1.4}$$

where, $\mathbf{X}$ is the colour vector for a given pixel and $\theta \in [0, 1]$. Now, from the Bayes law, a posteriori probability as given in eq. 1.4 can be expressed as,

$$P\left(Skin|\mathbf{X}\right) = \frac{P(Skin)P(\mathbf{X}|Skin)}{P(Skin)P(\mathbf{X}|Skin) + P(Non-skin)P(\mathbf{X}|Non-skin)} \tag{1.5}$$

where, $P(\mathbf{X}|Skin)$ and $P(\mathbf{X}|Non-skin)$ are skin and non-skin likelihoods for pixels respectively, and $P(Skin)$ and $P(Non-skin)$ are the corresponding priors. For a given set of skin and non-skin histograms, the skin and non-skin likelihoods can be obtained as:

$$P(\mathbf{X}|Skin) = \frac{s\left(\mathbf{X}\right)}{T_s}, P(\mathbf{X}|Non-skin) = \frac{n\left(\mathbf{X}\right)}{T_n} \tag{1.6}$$

where, $s\left(\mathbf{X}\right)$ and $n\left(\mathbf{X}\right)$ are pixel counts in the colour $\mathbf{X}$-bin of the skin and non-skin histograms, respectively. T_s and T_n are the total counts in skin and non-skin histograms. Therefore, the skin prior $P(Skin)$ and the non-skin prior $P(Non-skin)$ can be approximated as:

$$P(Skin) = \frac{T_s}{T_s + T_n}, P(Non-skin) = 1 - P(Skin) \tag{1.7}$$

Some of the state-of-the-art skin detection methods which use Bayes classifier are given in [13, 34, 37, 50, 51]

1.2.2 Gaussian classifiers

In recent years, many of the skin detection methods used a Gaussian or a mixture of Gaussians for skin distribution modelling. The major advantage of these parametric models is that they require less training data for generalization, and thus have lesser storage requirements.

1.2.2.1 Single Gaussian model

Under certain illuminating conditions, the skin colour distribution of different individuals belonging to a particular race can be modelled by using a multivariate single Gaussian distribution function [52, 53]. The skin-colour distribution is expressed as:

$$P(\mathbf{X}) = \frac{1}{\sqrt{(2\pi)^d |\mathbf{\Sigma}|}} \exp\left[-\frac{1}{2}(\mathbf{X}-\mu)^T \mathbf{\Sigma}^{-1}(\mathbf{X}-\mu)\right] \tag{1.8}$$

where, μ and $\mathbf{\Sigma}$ are the mean vector and the covariance matrix respectively; d is the colour space dimension. The pixel classification is done by comparing the probability $P(\mathbf{X})$ with a given threshold value. The threshold value can be obtained from the ROC, which is obtained using a set of training data.

1.2.2.2 Gaussian mixture model

Human skin colour shows significant chromatic variation among different races. So, multiple modes can co-exist within the skin colour cluster of different people from different races, and thus the colour distribution cannot be modelled efficiently using single Gaussian distribution function [54]. Also, illumination variation causes significant deviation of skin colour appearance. This can generate multiple modes in skin colour distribution, and thus single mode assumption of skin colour distribution does not hold. Hence, many researchers used Gaussian mixture models (GMMs) for skin colour distribution modelling to describe complexly shaped distributions. The skin colour distribution can be modelled by using a Gaussian mixture model with K number of Gaussian components as:

$$P(\mathbf{X}|\Psi) = \sum_{i=1}^{K} \omega_i \times P(\mathbf{X}|i;\psi_i) \tag{1.9}$$

where

$$P\left(\mathbf{X}\left|i;\psi_i\right.\right)=\frac{1}{\sqrt{(2\pi)^d\left|\Sigma_i\right|}}\exp\left[-\frac{1}{2}(\mathbf{X}-\mu_i)^T\Sigma_i^{-1}\left(\mathbf{X}-\mu_i\right)\right] \tag{1.10}$$

In this, the parameters μ_i, $\mathbf{\Sigma}_i$, ω_i are the mean, the covariance matrix, and the prior probability weight for the i^{th} Gaussian, respectively; $\Psi=(\psi_1,..,\psi_K)$ is the set of parameter vectors, where $\psi_i=\{\mu_i,\mathbf{\Sigma}_i\}$. All these parameter values except K are obtained by the Estimation-Maximization (EM) algorithm.

The EM algorithm iteratively finds the maximum likelihood (ML) function:

$$L(\mathbb{X};\Psi)=\prod_{n=1}^{N}P\left(\mathbf{X}_n\left|\psi_i\right.\right) \tag{1.11}$$

Here, $\mathbb{X}=\{\mathbf{X}_1,..,\mathbf{X}_N\}$ is a set of sample vectors $\mathbf{X}_i$. The above function is used to estimate the model parameters, which best describe the distribution of data. The EM algorithm has two steps, which are:

- **Expectation step:**

  In this step, the expected value of the log-likelihood function is calculated as:

$$Q\left(\Psi\left|\Psi^{(t)}\right.\right)=\mathrm{E}\left[\log L(\mathbb{X};\Psi)\left|\mathbb{X},\Psi^{(t)}\right.\right] \tag{1.12}$$

  where, $\Psi^{(t)}$ is the parameter set time t.

- **Maximization step:**

  In this step, parameters are calculated that maximize the log-likelihood as:

$$\Psi^{(t+1)}=\arg\max_{\Psi}Q\left(\Psi\left|\Psi^{(t)}\right.\right) \tag{1.13}$$

  Subsequently, the parameters of GMM are obtained as:

$$\hat{\mu}_i^{(t+1)}=\frac{\sum_{n=1}^{N}\mathbf{X}_nP^{(t)}\left(i\left|\mathbf{X}_n\right.\right)}{\sum_{n=1}^{N}P^{(t)}\left(i\left|\mathbf{X}_n\right.\right)}\quad, \tag{1.14}$$

$$\hat{\mathbf{\Sigma}}_i^{(t+1)} = \frac{\sum_{n=1}^{N} \left(\mathbf{X}_n - \hat{\mu}_i\right)\left(\mathbf{X}_n - \hat{\mu}_i\right)^T P^{(t)}\left(i\,|\mathbf{X}_n\right)}{\sum_{n=1}^{N} P^{(t)}\left(i\,|\mathbf{X}_n\right)} \quad , \tag{1.15}$$

$$\hat{\omega}_i^{(t+1)} = \frac{1}{N}\sum_{n=1}^{N} P^{(t)}\left(i\,|\mathbf{X}_n\right) \quad , \tag{1.16}$$

$$P^{(t)}\left(i\,|\mathbf{X}_n\right) = \frac{\hat{\omega}_i^{(t)} P\left(\mathbf{X}_n \left| \psi_i^{(t)}\right.\right)}{P\left(\mathbf{X}_n\,|\Psi^{(t)}\right)} \quad . \tag{1.17}$$

For proper convergence, a good initial specification of K and other parameters is needed. A k-Means clustering of the training data can provide the initial parameter values for the EM algorithm. The skin classification process is same as described for single Gaussian modelling.

1.2.3 Elliptical boundary model

The major limitation of GMMs is that they are computationally expensive. To reduce the computational burden, Lee and Yoo [55] proposed an "elliptical boundary model" as an alternative to GMMs. In some restricted situations, it produces performance comparable to that of GMMs with much lesser computational cost. The elliptical boundary model $\Phi = (\mathbf{X}; \psi, \Lambda)$ is defined as:

$$\Phi\left(\mathbf{X}\right) = \left[\mathbf{X} - \psi\right]^T \Lambda^{-1} \left[\mathbf{X} - \psi\right] \tag{1.18}$$

where,

$$\psi = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \quad , \tag{1.19}$$

$$\Lambda = \frac{1}{N}\sum_{i=1}^{n} f_i\left(\mathbf{X}_i - \mu\right)\left(\mathbf{X}_i - \mu\right)^T \quad , \tag{1.20}$$

In this, $\mathbf{X}_i$ is the i^{th} distinctive chrominance vector and f_i is the number of samples having chrominance vector $\mathbf{X}_i$; n is the number of distinctive coloured pixels in training data. Here, $N = \sum_{i=1}^{n} f_i$ and $\mu = \frac{1}{N}\sum_{i=1}^{n} f_i \mathbf{X}_i$.

For classification, the value $\Phi\left(\mathbf{X}\right)$ is compared with a threshold θ. If $\Phi\left(\mathbf{X}\right) \geq \theta$ the corresponding pixel is classified as skin, and non-skin otherwise.

1.2.4 Artificial neural network

An ANN is a biologically inspired statistical learning algorithm for functional approximation, pattern recognition, and classification. ANNs can be used as a supervised classifier for skin classification. Training is performed using a set of labelled input skin and non-skin pixels in a given colour space. The ANN classifies new input patterns within the labelled classes. In a feed-forward ANN, an iterative weight update is performed through a gradient descent algorithm, also termed as the back-propagation algorithm. In the training phase, the network processes its inputs in a feed-forward fashion. Subsequently, resulting outputs are compared with their expected values, and resultant errors are back-propagated to update the weights according to their contributions to the overall error. The performance of an ANN is primarily dependent on two factors – number of hidden layers and number of hidden nodes. Therefore, an extensive tuning is needed for an ANN to get its optimal performance. In skin detection, the class conditional distributions for skin and non-skin pixels are learned by the neural networks during the training. Some of the skin detection methods which use ANN are given in [56–59].

1.2.5 Support vector machine

Some researchers, such as Casiraghi *et al.* [60], Schettini *et al.* [61] used Support Vector Machine (SVM) classifier [62] for skin classification. SVMs were originally designed for two-class classification, and thus it is suitable for skin and non-skin classification. If data are linearly separable, SVM finds a hyperplane which separates two classes with maximum marginal distance. This method non-linearly maps the input data (if not linearly separable in current feature space) to some higher dimensional space, where the data can be linearly separated. This mapping from lower to higher dimensional spaces makes the classification simpler and more accurate.

1.2.6 Random forest

Tree-based classifiers are popular due to their intuitive appeal and easy training procedures. Tin Ho [63] introduced a Random Forest architecture. The random forest is a collection of tree predictors where a random vector governs each of them. The random vector is obtained by independent sampling of the same distribution for each of the trees [64]. For an object

classification, each of the trees is fed with an input vector, and the tree gives a "vote" to it for a class. The majority of votes decides the class of the input vector. Khan *et al.* [17] showed that skin detection can also be possible with the help of a random forest approach.

1.3 Evaluation Procedure

1.3.1 Evaluation metrics

The performance of a skin detection method can be measured based on four primary metrics: number of truly detected skin pixels (*i.e.*, true positives $= TP$); number of truly detected non-skin pixels (*i.e.*, true positives $= TN$); number of falsely detected skin pixels as non-skin pixels (*i.e.*, false negatives $= FN$); and number of falsely detected non-skin pixels as skin pixels (*i.e.*, false positives $= FN$). From these values, following measures can be computed:

- *False positive rate:* $\delta_{fp} = \frac{FP}{FP+TN}$ *i.e.*, the percentage of non-skin pixels falsely classified as skin [9].
- *False negative rate:* $\delta_{fn} = \frac{FN}{FN+TP}$ *i.e.*, the percentage of skin pixels falsely classified as non-skin [9].
- *Total detection rate:* $\delta_t = \frac{FP}{FP+TN} + \frac{FN}{FN+TP}$ [9].
- *Accuracy* $= \frac{TP+TN}{TP+TN+FP+FN}$ [22].

In case of non-binary classification, where a pixel has a skin probability, the false positive and false negative error rates are calculated based on a threshold. The false negative rate decreases with high threshold values. However, this increases the false positive rate. These two errors can be related with the help of a receiver operating characteristics (ROC) curve [65]. The area under the ROC curve is used as a measure of effectiveness of a skin detection method [66].

1.3.2 Datasets

There are number of skin datasets for skin segmentation. One of the most frequently used dataset is ECU face and skin detection dataset [65]. The ECU dataset contains 2 sets (each contains 2000 images) of skin annotated images. The images in ECU dataset are captured in uncontrolled indoor and outdoor environments. Out of this two sets, one is generally used

for training, whereas the other is used for validation. Kawulok et al. [9, 14, 67] introduced another dataset for skin detection, which is named as *hand gesture recognition dataset* or HGR dataset. The HGR dataset contains three sets of skin annotated images, out of which the first set contains 899 images, while the other two sets contain 85 and 574 images respectively. All the images in this dataset are captured in an indoor but uncontrolled environments. Images belonging to the first set (899 images) are captured under non-uniform illumination in the presence of skin-like background colours for some of the images. The second set contains 85 high resolution images of hand poses captured in an environment similar to the first set. The third set contains high resolution images of hand poses, captured under varying illumination condtions in presense of a green/blue background. Apart from these two datasets, another two datasets namely Cambridge hand gesture dataset [68, 69], NUS dataset [70] are used for validation of the proposed methods.

1.4 Overview of Different Methods of Skin Detection

Most of the existing skin detection methods can be broadly classified according to their applicabilities and approaches, such as static framework-based methods and dynamic framework-based methods as shown in Figure 1.5. Skin detection methods which use static frameworks can be used for both images and videos with static environments. Methods which use static frameworks can be further classified into three groups based on the modelling of skin colour distribution, such as explicit boundary specification, parametric modelling and non-parametric modelling. On the other hand, dynamic framework-based methods are used for videos having dynamic environments, such as varying illumination and dynamic background conditions. A more detailed discussion on existing skin detection methods is given as below.

1.4.1 Skin detection methods using static framework

Majority of existing skin detection methods use a static framework. This implies that characteristics of background and illumination do not vary with time. Hence, these type of skin detection methods are intended for both images and videos for static environments. These methods either use a global skin detection model or a local skin detection model. Kakumanu *et al.* [71] and Kawulok *et al.* [8] provided comprehensive surveys of different approaches of skin

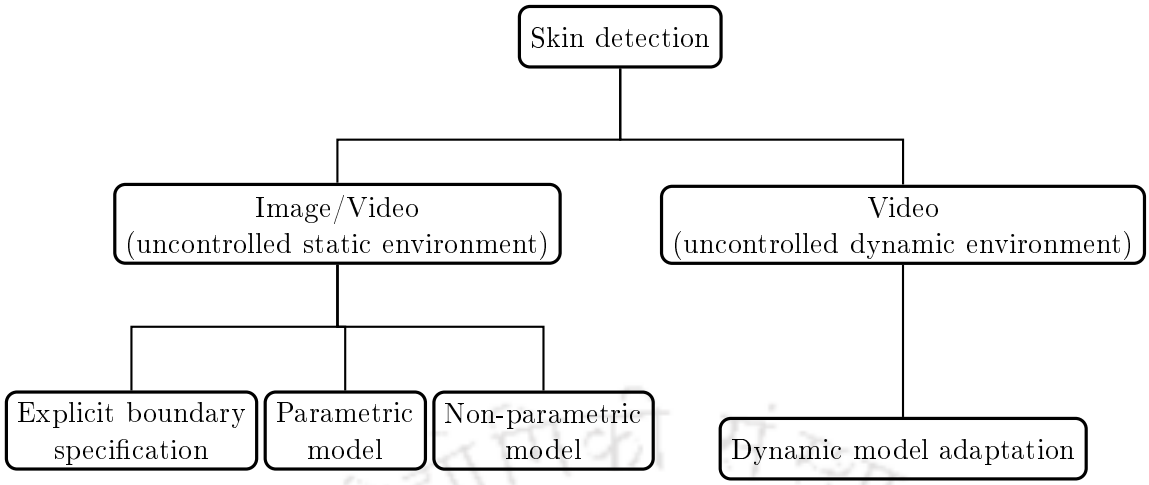


Figure 1.5: Skin detection methods

colour modelling and classification. Skin segmentation methods can be broadly classified into three major classes: *Explicit Boundary Specification*, *Parametric Modelling-based approach*, and *Non-Parametric Modelling-based approach*.

1.4.1.1 Explicit boundary specification

Boundary specification for skin colour depends on a set of thresholds and conditions which could be either defined in the same colour space, (*e.g.* RGB) or in a transformed colour space, such as YCbCr, HSV, CIELab etc.

One of the earliest methods of skin detection is proposed by Sobottka and Pitas [43]. They proposed a skin detection boundary along S and H channels in HSV colour space as $S \in [0.23, 0.68]$ and $H \in [0, 50]$. Later, Tsekeridou and Pitas proposed a modification [2] to this method for face region segmentation in an image watermarking system [72]. The corresponding boundary rule in the HSV colour space is as follows:

$$\begin{cases} (0 \leq H \leqslant 25) \vee (335 \ \leqslant \ H \leqslant 360) \\ (0 \leqslant S \leqslant 0.6) \wedge (0.4 \leqslant V) \end{cases} \tag{1.21}$$

Figure 1.6-a shows the projection of the above rules onto the RGB colour space. Here, darker shade shade implies higher density of skin pixels. Solina et al. [3] proposed another set of fixed

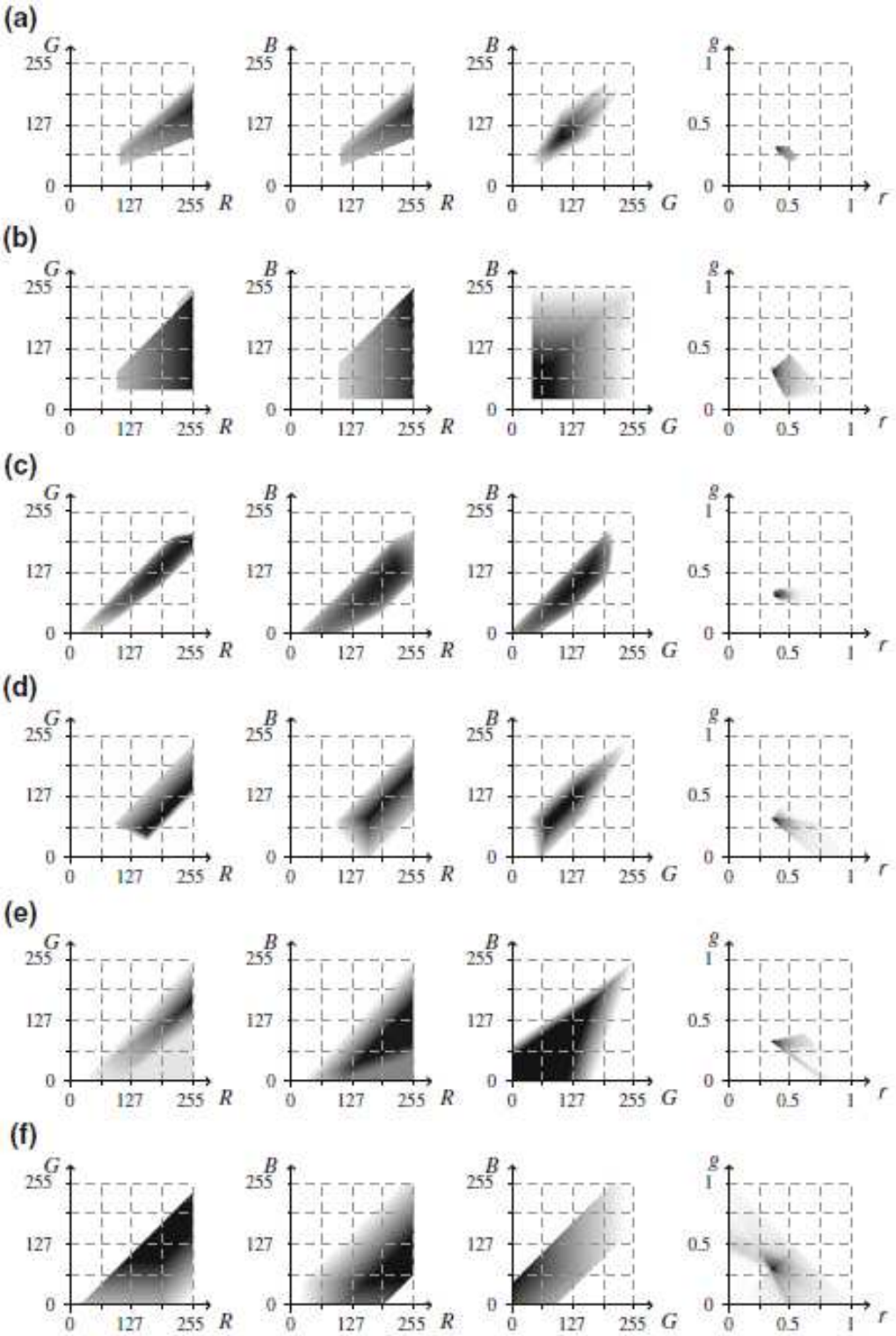


Figure 1.6: Skin colour distributions in different colour planes: (a) Tsekeridou and Pitas [2], (b) Solina et al. [3], (c) Hsu et al. [4], (d)Kukharev and Nowosielski [5], (e) A. Cheddad et al. [6], and (f) Y.-H. Chen et al. [7]. (Note: the figure is taken from [8])

rules in RGB space for face detection of people having fair complexion as:

$$\begin{cases} (R > 95) \wedge (G > 40) \wedge (B > 20) \\ \max(R,G,B) - \min(R,G,B) > 15 \quad \text{in uniform daylight illumination} \\ |R - G| \leqslant 15 \wedge (R > G) \wedge (R > B) \end{cases} \tag{1.22}$$

or,

$$\begin{cases} (R > 220) \wedge (G > 210) \wedge (B > 170) \\ |R - G| \leqslant 15 \wedge (R > G) \wedge (R > B) \end{cases} \text{in flashlight lateral illumination} \tag{1.23}$$

For unknown lighting conditions, a pixel is classified as skin if it satisfies one of the above two conditions. These rules are illustrated in Figure 1.6-b in R-G, R-B, G-B and r-g planes. Hsu et al. [4] proposed a boundary rule based on YCbCr colour space. The authors observed that the shape of skin tone cluster in Cb-Cr space can be approximated as an elliptical structure where the cluster location depends on luminance Y. They performed a non-linear modification to C_b and C_r values if $Y < 125$ or $Y > 188$. Subsequently, the skin pixel cluster is modelled as an ellipse in a transformed space $Cb'Cr'$. The equivalent results for skin distribution in RGB space is shown in Figure 1.6-c. Kukharev and Nowosielski proposed another set of skin detection rules [5] using RGB and YCbCr colour spaces as follows:

$$\begin{cases} (R > G) \wedge (R > B) \\ \{(G \geqslant B) \wedge (5R - 12G + 7B \geqslant 0)\} \vee \{(G < B) \wedge (5R + 7G - 12B \geqslant 0)\} \\ \{Cr \in (135, 180)\} \wedge \{Cb \in (85, 135)\} \wedge (Y > 80) \end{cases} \tag{1.24}$$

The corresponding model representation in RGB and rg space is given in Figure 1.6-d. Cheddad *et al.* [6] transformed the normalized RGB colour space into a single-dimensional error signal, where the skin colour distribution can be modelled as a Gaussian curve [6]. Subsequently, a pixel is classified as a skin if its 1D equivalent value lies within the two threshold values determined by the standard deviation of the curve. The skin model is shown in Figure 1.6-e. Recently, Chen et al. proposed a new RGB subspace for skin detection by subtracting the RGB values: $sR = R - G$, $sG = G - B$, $sB = R - B$. Subsequently, they proposed a boundary rule $\{(-142 < sR < 18) \wedge (-48 < sG < 92) \wedge (-32 < sB < 192)\}$. The rules are illustrated in Figure 1.6-f. In [73], Shaik *et al.* compared HSV and YCbCr spaces for skin detection using a

Table 1.1: Most popular examples of colour spaces used in skin detection: RGB, YCbCr, HSV and advanced space ER/GH [18]

Colour space	Range of components	Restrictions for skin colour
RGB	R, G, B: [0, 255]	$R > 95 \wedge G > 40 \wedge B > 20 \wedge \{\max(R,G,B) - \min(R,G,B) > 15\} \wedge \lvert R - G \rvert > 15 \wedge R > G \wedge R > B$
YCbCr	Y, Cb, Cr: [0, 255]	$Y > 80 \wedge 77 < Cb < 127 \wedge 133 < Cr < 173$
HSV	H: [0°, 360°], S, V: [0, 1]	$0° < H < 50° \wedge 0.1 < S < 0.68 \wedge 0.35 < V < 1$
ER/GH	R, G: [0, 255], H: [0°, 360°]	$13.4224 < E \wedge R/G < 1.7602 \wedge H < 23.89$

boundary-based method. In 2015, Sawicki and Miziolek [18] proposed another set of boundary rules in CMYK space as follows:

- Before ROC analysis:

$$(K < 205) \wedge (0 \leqslant C \leqslant 0.05) \wedge (0.089 < Y < 1) \wedge (0 \leqslant C/Y < 1) \wedge (0.1 \leqslant Y/M < 4.8) \tag{1.25}$$

- After ROC analysis:

$$(K < 205) \wedge (0 \leqslant C \leqslant 0.05) \wedge (0.0909 < Y < 0.945) \wedge (0.1 \leqslant Y/M < 4.67) \tag{1.26}$$

Apart from these simple boundary specifications in different colour spaces, advanced approaches are also proposed for a more accurate 3D description of skin cluster. For example, Garcia and Tziritas [74] proposed a skin detection method by utilizing a set of planes in the YCbCr space. Brand and Mason [50] performed a comparative analysis of algorithms in three colour spaces: RGB, YES and YIQ. In their analysis, parametric thresholds and statistical functions are used. Thresholding of the R/G ratios is also performed. In [51], a new colour space ER/GH is proposed by mixing of colour components. In the pseudo space ER/GH, E belongs to YES, the R/G ratio is from RGB space, and H is from HSV.

Some of the authors used additional information like texture features to improve skin detection. For example, Wang *et al.* [75] used gray-level co-occurrence matrix (GLCM) for skin detection. In this method, a white balancing is performed in YCbCr colour space to minimize

the effect of uncontrolled illumination conditions. Firstly, the Y components are arranged in descending order. The minimum value of the top 5% values of the Y component is termed as a parameter E, and remaining values in the top 5% are set to 255. Similarly, the maximum value among the bottom 5% values of the Y component is termed as a parameter B, and remaining values in the bottom 5% are set to 0. Finally, the intermediate Y components are re-calculated as:

$$g(x,y) = 255 \times \frac{\ln f(x,y) - \ln B}{\ln E - \ln B} \tag{1.27}$$

where, $g(x,y)$ is the white balanced luminance value at location (x,y), and $f(x,y)$ is the original luminance value before white balancing. The skin colour model is defined by a set of boundary rules in RGB space. The authors also found that skin distribution in YCgCb space takes circular shape. Finally, a skin mask is obtained by ANDing two skin models derived from RGB and YCbCr spaces. Detection performance is further improved by incorporating a texture analysis into this skin model. Textural features are extracted using the GLCM. For a given gray-scale image I of size $n \times m$, the GLCM is given by:

$$T(i,j) = \sum_{x=1}^{n}\sum_{y=1}^{m} \begin{cases} 1, \text{if} I(x,y) = i \wedge I(x+\Delta_x, y+\Delta_y) = j \\ 0, \text{otherwise}, \end{cases} \tag{1.28}$$

where, (Δ_x, Δ_y) is the offset between the pixels $I(x,y)$ and $I(x+\Delta_x, y+\Delta_y)$. The computational complexity in determining the GLCM depends on the number of grey levels g, and it is proportional to $O(g^2)$

However, recently published literatures show that the performance of explicit boundary specification-based methods are not better than the model-based approaches [8].

1.4.1.2 Parametric modelling

In order to improve detection accuracy, some researchers used Parametric Modelling-based approaches for skin detection. For example, Yang *et al.* [53] used a single multivariate Gaussian to model skin colour distribution. But, skin colour distribution possesses multiple co-existing modes. So, a Gaussian Mixture Model (GMM) ([54]) is more appropriate than a single Gaussian function. Selection of the number of Gaussian components (K) in GMM is very important for proper modelling of pixel distribution. Different researchers used different values of K in their methods. Yang and Ahuja [76] used two Gaussian components in CIE-Luv colour space to model

skin colour distribution. In this work, statistical tests are provided to showcase the advantage of using GMM over SGM for skin colour distribution modelling. Greenspan et al. [77] showed that GMM-based representation of skin pixel distribution is more robust to environmental changes, such as colour space changes, highlights and shadows. They also used two Gaussian components for GMM, and one represents the distribution of skin colour under normal light, while the other represents the distribution of the more highlighted regions of the skin. Caetano et al. [78] used two to eight Gaussian components for pixel distribution modelling in rg colour space for people having different skin tones. Lee and Yoo [55] proposed an elliptical modelling-based approach for skin detection. The elliptical modelling is less computationally complex than the GMM modelling. However, many true skin pixels may be rejected if the ellipse is small. On the other hand, if the ellipse is sufficiently large, many non-skin pixels may be detected as skin pixels. They used six Gaussian components to implement the GMM. On the other hand, Thu *et al.* [79] used four Gaussian components. Use of multiple Gaussian enables detection of different parts of a face which are illuminated differently. Jones and Rehg [13] used two separate GMMs, each having 16 Gaussian components for skin and non-skin pixel distribution. A skin probability map (SPM) for an image is derived from the two models using Bayes theorem. The SPM is a 2D array of size equal to the image. An element of the SPM represents a posteriori probability of a pixel being skin at that location.

The performance of these simple parametric models is limited due to two major factors – a) apparent change in skin appearance due to uncontrolled illumination conditions, and b) the presence of skin-like colours in image background. To overcome these problems, different authors proposed different improvements over simple parametric models for skin detection. Phung *et al.* [38] proposed an adaptive scheme to select the optimum threshold for the SPM by assuming that a skin region to be coherent and homogeneous in texture. Segmentation accuracy of skin regions can be further improved by incorporating texture analysis in the parametric modelling framework. Texture features can be extracted by performing texture analysis in various domains, such as grayscale [75, 80], colour [81], or skin map [82]. In order to extract texture features, different authors used different feature descriptors. Jiang et al. [83] proposed a new approach by incorporating texture and space analysis in a standard SPM framework. An initial skin mask is derived for an image by thresholding the SPM with a low threshold. Subsequently, textural features are extracted using Gabor wavelets. This gives a

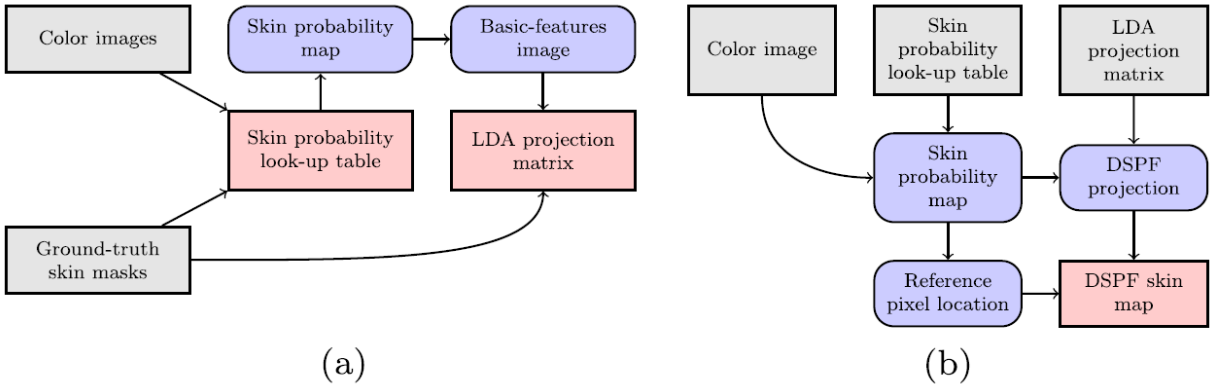


Figure 1.7: A flowchart showing (a) training and (b) detection processes proposed by Kawulok [9]

textural map for the image. The texture map is thresholded based on an assumption that background regions are coarser than skin regions. This gives a texture mask or a texture filter which is later combined with the initial skin mask to obtain a more accurate skin mask. This reduces false acceptance error significantly. Finally, the watershed segmentation is employed with a set of well-defined region markers to grow skin regions to reduce false rejection error. H.-M. Sun [84] proposed a local adapation scheme for the Bayesian classifier as proposed by Jones and Rehg [13]. They generated a local skin model from a set of skin pixels samples from the image. Finally, the local model is combined with the global or trained model in a weighted sum approach. P. Ng and C.M. Pun combined 2-D Daubechies wavelets -based texture analysis with a GMM-based colour model [85]. The 2-D Daubechies wavelets are calculated by using the sub-images which are centered at each of the pixel locations. Texture feature at each pixel location is represented by the wavelet energy vector $\mathbf{v_e}$, which is obtained by applying Shannon entropy on the wavelet coefficients vector $\mathbf{v_c}$. The $\mathbf{v_e}$ for all the pixel locations are finally grouped into a set of clusters using k-Means clustering algorithm. Finally, some of the clusters are marked as non-skin based on their Shanon entropies and eliminated accordingly. Kawulok *et al.* [9] used linear discriminative analysis (LDA) to derive discriminative features between skin and non-skin regions. In this method, LDA projection matrix is derived by using colour and local texture features from a set of labelled images. The LDA projection matrix depends on training data. Therefore, LDA gives a projection matrix which ensures best possible inter-class discrimination.

Another approach follows an use of spatial analysis of skin regions by exploiting the spatial alignment of skin pixels and their relation with neighbourhood pixels [14, 80, 86, 87]. These

approaches significantly reduce false positives in detecting the skin regions. In general, all these spatial analysis-based methods are based on a standard SPM. Ruiz-del-Solar and Verschae [86] proposed a skin detection method which uses a controlled diffusion. The controlled diffusion process has two steps: a) extraction of diffusion seeds, and b) actual diffusion process. The diffusion seeds are extracted by thresholding the SPM with a high threshold. In the diffusion step, skin regions are grown from the seeds by including the neighborhood pixels which satisfy a given diffusion criteria. The criteria depends on two factors – a) difference between source and a test pixel in diffusion domain, and b) SPM value at the test pixel location. Therefore, this method works well if skin regions have sharp boundaries. A leak in diffusion may occur if there are smooth transitions between pixels from one region to another. In 2010, Kawulok proposed an energy-based scheme for skin blob analysis [87]. Pixels with high valued SPM values are selected as skin seeds. These seed regions are subjected to morphological erosion to further reduce false acceptance. In this method, seed pixels are assumed to have a maximal energy, which is likely to be spread over an image. The amount of energy transferred to an adjacent pixel from a source pixel depends on the skin probability of the adjacent pixel. A pixel is excluded from skin region if there is no energy left to be passed onto it from a source pixel. In 2013, M. Kawulok [14] proposed a propagation-based region growing method, which utilises spatial relationship between the pixels. Kawulok's method is based on Dijkstra's minimum path-cost algorithm [88]. In Kawulok's method, each pixel is considered as an independent node and the image is the corresponding graph. In this method, the optimum values of region growing parameters are selected manually.

There are another class of approaches of skin segmentation, which use some prior information about the actual skin colour of a person present in an image. In general, human skin colour does not show significant variations over the body. So, a face detector can be used to detect the face and extract a set of pixels beloning to facial region. The prior information obtained from the facial pixels is then utilized to segment out other skin regions of the human body. A global skin detection model can be locally adapted according to the distribution of facial skin pixels. Fritsch *et al.* [89] used face detection to derive a local skin model for skin region tracking. In 2008, Kawulok [19] proposed a dynamic skin model by using pixel characteristics of facial regions. The global pixel statistics are fused with the local statsitics of facial skin pixels. Yogarajah *et al.* [20] used a dynamic thresholding-based method for skin detection. In

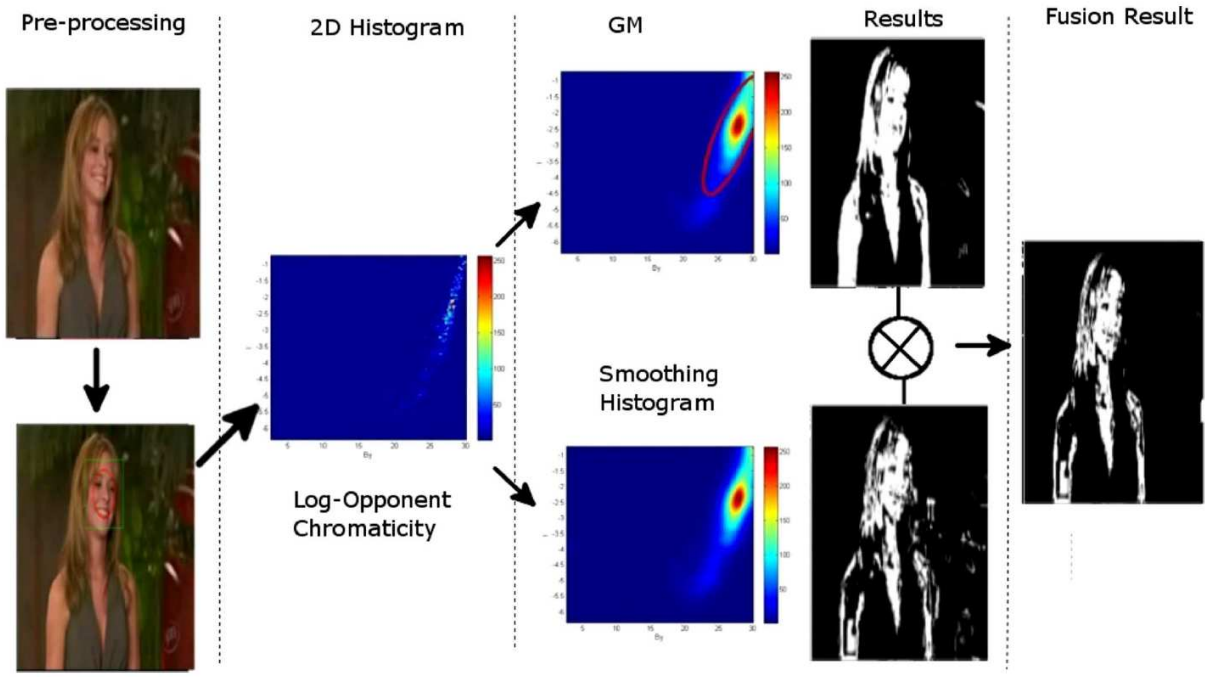


Figure 1.8: Proposed framework by Tan *et al.* [10]: eye detector, 2-D histogram, Gaussian model, and fusion strategy.

this method, a dynamic threshold is obtained from the characteristics of skin pixels extracted from the facial regions. Tan *et al.* [10] proposed a fusion-based skin detecion method using face detection. For this, a smoothed colour histogram and a Gaussian model of skin is fused together. Kawulok *et al.* [21] showed that the selection of seeds points using facial pixels can improve the detection accuracy in a region growing-based skin detection method. Pixels extracted from the face provide a good estimate of colour distribution of skin regions even in the presence of skin-like backgrounds and/or poor illumination conditions.

1.4.1.3 Non-parametric modelling

Skin regions can also be detected using non-parametric modelling-based approaches without deriving any explicit model (*e.g.*, SGM or GMM) for skin colour distribution. The simplest non-parametric classifier estimates skin colour distribution from a colour histogram derived from a set of skin pixel samples [90–93]. The major advantage of these approaches is that they there is no requirement of shape estimation of skin colour distribution. However, these approaches require a very large amount of training samples to properly model the skin colour distribution. Therefore, these methods are applicable to a limited range of imaging conditions.

Cortes and Vapnic showed that Support vector machines (SVMs) can also be used for skin detection. In general, the number training samples for skin and non-skin pixels usually become too large to handle by the SVM. Han *et al.* [25] proposed a skin segmentation method using active learning based SVM classifier and region information. The SVM active learning is a well-known approach to deal with large training dataset [94]. In this method, it assumed that the region information is robust to illumination variations and noise. Each image is divided into a number of regions. A region is selected as a skin if it satisfies the following criterion, which is expressed as:

$$\frac{NS(R_i)}{NT(R_i)} > \eta \tag{1.29}$$

where, $NS(R_i)$, $NT(R_i)$ are the number of skin pixels and the total number of pixels in the region R_i, respectively; η is a pre-defined constant.

A more popular non-parametric approach is the use of back propagation ANNs (BPANNs) for skin detection [11, 95]. For example, Chen *et al.* [58] proposed a skin detection algorithm by using BPANN with genetic optimization. In their work, pixel components in RGB space are transformed into the normalized RGB space. The r and g components of pixels then fed into a BPANN made of 2 input neurons, 4 hidden neurons in two hidden layers, and an output neuron. Each of these neuron's response is characterised by a logistic sigmoid function given by:

$$f(x) = \frac{1}{1 + e^{\sigma x}} \tag{1.30}$$

where, σ is the steepness of the sigmoid curve, x is the weighted sum of the inputs, and $f(x)$ is the output. The stability and convergence of the ANN depends on the parameter σ. So, they used a genetic algorithm (GA) to optimize the selection of the parameter σ. Finally, if the colour components r, g, b in normalized RGB space satisfy $r > g$ or $r > b$, then the r and g components are fed into the BPANN classifier. Seow *et al.* [11] proposed a skin colour model for face detection, which aims at reducing the effect of skin colour variations among different people. They used a 3-layered BPANN with the r, g, b components as inputs as shown in Figure 1.9. A set of 410 skin samples (each containing a 10×10 patch) is collected from skin regions belonging to different races. Since the sample set cannot represent the entire skin population, a Multi-Layer Perceptron (MLP) ANN is trained by using a Back Propagation (BP) algorithm for interpolation of sample set. Finally, a $256 \times 256 \times 256$ colour cube is generated to obtain all the

possible colour combinations and they are fed into the MLP to extract the skin regions. Yang

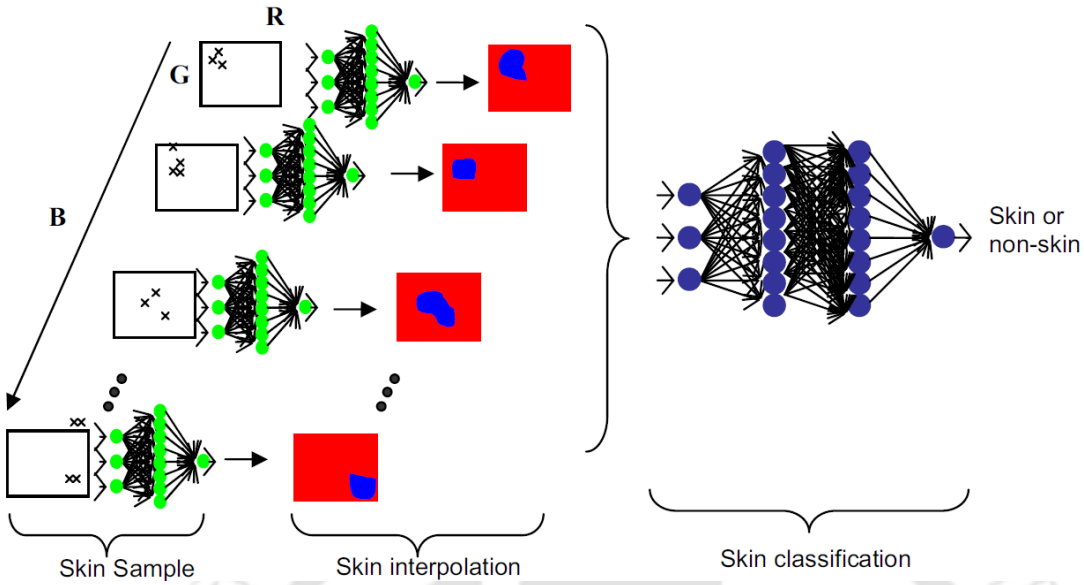


Figure 1.9: Skin detection using ANN proposed by Seow *et al.* [11]

et al. used ANN along with an adaptive skin modelling to detect skin regions more accurately in an image as shown in Figure 1.10.In this method, the luminance component Y of the YCbCr colour space is used for reducing the effects of illumination variations. At first, the Y component is arranged in descending order, and divided into multiple equidistant intervals. After that, pixels belonging to same luminance interval are selected, and the corresponding mean and covariance matrix of the selected pixels in Cb, Cr space are calculated. The luminance mean of each interval, the covariance and mean in Cb, Cr space are used to train a three layer ANN. Finally, the output of the ANN is fed to a Gaussian classier for skin classification.

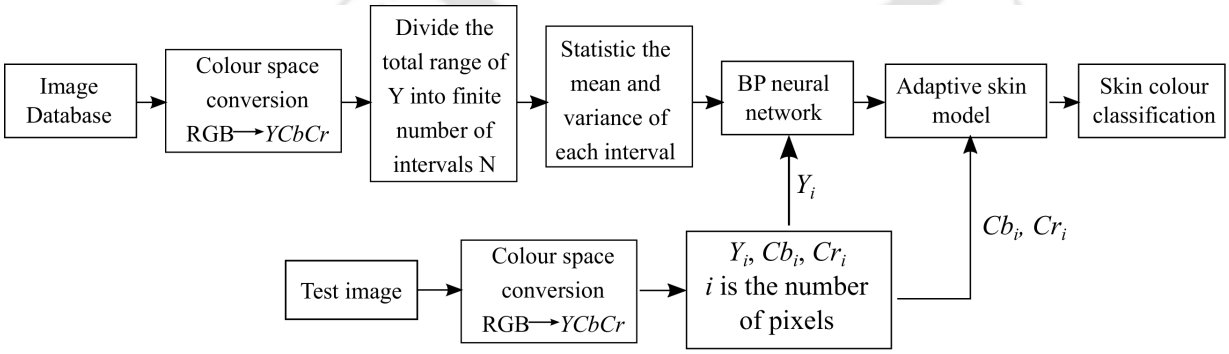


Figure 1.10: Proposed framework by Yang *et al.* [12].

1.4.2 Skin detection methods using dynamic architecture

In HCI, detection of human body parts is an important step. However, the task is not so easy as there are many constraints in locating the skin coloured regions in the videos. One of the major challenges is the effect of illumination change on skin colour appearance. Varying illumination can make skin regions appeared to be different from its actual colour. Illumination variation can occur in two ways - globally and locally. A global illumination variation occurs when the characteristics of the illuminating source change with time. On the other hand, local illumination variation occurs when illumination becomes non-uniform over the exposed skin region regions. One of the major reason behind the non-uniform illumination is the curved nature of skin surface. This causes the formation of form shadows on the skin surface. Another reason is the formation of cast shadows produced by one moving body part onto another body part. In case of directional light sources, amount of illumination on a skin patch also depends on its orientation with respect to the light source. It is found that the occurrence of local illumination variation is more frequent as compared to global illumination change for most of the HCI applications.

In recent years, many research works [9, 16, 22, 23, 96] have been reported on skin detection under unconstrained environments for images. However, only few research works have dealt with skin detection in videos under varying illumination conditions. Soriano *et al.* [97] investigated the effects of static but non-uniform illumination on skin colour appearance. In their work, a "skin locus" is introduced to describe a chromatic constraint on skin colour apperance under non-uniform lighting conditions. However, derivation of such a skin locus requires a calibrated camera and the face has to be captured for different illuminants. Also, the skin locus is camera specific, and hence it has to be calculated for every unknown videos. Sigal *et al.* [98] proposed a dynamic skin model for varying illumination conditions. In this method, a dynamic histogram adaption model based on a second order Markov model is proposed. In this method, it is considered that the illumination changes gradually and globally. Therefore, this approach is not suitable for local illumination changes owing to their uncertain nature. Habili *et al.* [99] used motion information alongwith colour information to detect skin regions in a video. However, this method assumes that the background does not have skin-like coloured regions. Awad *et al.* [100] proposed a fusion-based model for skin detection in videos. However, this method assumes the absence of any skin-like colours in the background and uniform illumination of

skin regions. Therefore, under unconstrained illumination and background conditions, the skin pixels cannot be located perfectly. Also, it requires a set of labelled initial frames to train a Support Vector Machine (SVM) classifier, and the initial positions of skin-coloured objects need to be determined. Han *et al.* [101] proposed a skin segmentation and tracking algorithm for sign language recognition by using Support Vector Machine (SVM) active learning. The training of SVM is done by a set of initial frames. The active learning of SVM makes the algorithm computationally less expensive. However, the major drawback of this method is its inability to handle varying illumination conditions. The SVM needs to be re-learned at every frame to handle the varying illumination conditions. Training of an SVM repeatedly for every frame is a computationally very expensive process. Liu *et al.* [102] proposed a dynamic skin detection algorithm for videos. In this method, a face detection-based model update scheme is proposed for varying illumination conditions. However, only global illumination variations are considered, and the method is not suitable for local illumination changes which mostly occur due to moving body parts.

1.5 Research Motivation

From the brief literature survey presented in this report, it is evident that a significant amount of work is needed for efficiently detect skin regions in different environmental conditions in images. Additionally, detecting skin regions in videos in the presence of varying illumination conditions is another important task. To combine these requirements in one algorithm is a major challenge. Accordingly, this thesis looks into several aspects using chromatic and textural informations of skin regions and aims at developing suitable algorithms that take care of some limitations of the existing methods. The motivations behind this research work are given below:

(i) Typically a chromatic and/or textural discrimination is observed between skin and non-skin regions of an image. Kawulok *et al.* [9] used a linear discriminant analysis (LDA)-based most discriminative feature extraction approach for skin detection. However, extracted features totally depends on training data. As natural images are in general uncorrelated in a sense that spatial distribution of texel and colours is quite random. Hence, the discriminative features extracted by the LDA may not be most discriminative features for an unknown image. Hence, an image specific discriminative feature extraction

approach is necessary for the detection of skin regions in an unknown image.

(ii) It is observed that in spite of having colour deviation due to poor or non-uniform illumination, skin regions show more chromatic similarity with actual skin colours than the non-skin colours. Hence, it is necessary to identify and relate the skin regions, which look chromatically different due to improper illumination with the help of a set of actual skin samples obtained globally [13] or locally [10, 21].

(iii) Some of the recent research works [98, 102] consider global variation of background illumination without giving sufficient importance to local illumination variations in the frames of a video. Local illumination variation may occur in a video due to movement of hands or other body parts for some of the HCI applications, like gesture recognition. The local illumination variation changes the chromaticity of the skin regions. A static model-based skin detection method may not be able to identify the actual skin regions when the illumination changes locally. Therefore, it is necessary to follow an adaptive approach to update a skin model on the basis of local chromatic deformations of skin regions.

1.6 Objectives

The goal of this research work is to use chromatic and textural information of skin regions for detecting the skin regions in images and videos. For the detection of skin regions in images, two approaches can be followed: a) extraction of most discriminant features between skin and non-skin regions of an image, and b) determination of an image pixel distribution model (IDM) of an image. In the first approach, the most discriminative features between skin and non-skin regions need to be extracted from the test image itself. This would make the method more versatile and robust to scenic changes in images. In the second approach, a local skin model needs to be derived using the similarity between IDM and a reference skin pixel distribution model. The reference pixels can be extracted from the facial regions of the people present in the image. In case of videos, the skin model needs to be adaptive to local skin colour deformations due to local illumination variations. A skin detection method for an image has a static skin model, and thus, it may fail to detect skin regions in videos affected by the local illumination variations. The local illumination changes happen due to the movement of body parts. Hence, a skin distribution model for moving skin regions needs to derived, and this model can be

associated with the static skin model. The thesis also aims at additional reduction in detection errors by a region growing technique. **Chapter 1** of the thesis presents a brief literature review on different skin detection methods in images and videos, motivation and objectives of the thesis.

1.7 Thesis Organization

This thesis contains five chapters, including the present one.

Chapter 2 describes the proposed scheme for extracting discriminative features between skin and non-skin regions of an image. For this, a novel discriminative analysis method also termed as adaptive discriminative analysis (ADA) is proposed, which gives a discriminative feature space. Subsequently, a discriminative space map (DSM) is derived by using the discriminative feature space. A novel region growing method is also proposed, which grows skin regions of an image based on chromatic properties of an image. This results additional reduction in the detection errors.

In **Chapter 3**, image pixel distribution information is incorporated into a global skin detection model for a local model adaptation. To obtain a local skin detection model for an image, a local skin distribution model (LSDM) and a local background distribution model (LBDM) are needed, which can be derived by measuring similarity between an image pixel distribution model and a reference skin colour distribution model. The reference skin pixel distribution model can be derived from a set of skin samples of facial region of a person in an image. Finally, a novel skin map is obtained by performing a fusion of skin maps obtained from the global and the local skin detection models. The proposed skin map is termed as fusion-based skin probability map (FSPM), and finally the proposed dynamic region growing (DRG) algorithm is applied for additional detection error reduction.

Chapter 4 describes a skin detection method for videos when skin regions undergo chromatic appearance changes due to non-uniform illumination. Non-uniform illumination can take place due to local shading effects, which are associated with the motion of skin regions. The proposed method has two modules – a static skin detection model, and a dynamic skin detection model. The static model consists of a facial skin distribution model (FSDM) and a video specific background model. On the other hand, the dynamic model corresponds to a

moving skin-pixel distribution model (MSDM). The MSDM is dynamically updated based on associated chromatic changes in skin colour.

Finally, we draw our conclusion in **Chapter 5** by highlighting the strengths and shortcomings of our schemes and outlining possible extensions.

2

Detection of Skin using Image Specific Discriminative Feature Extraction

Accuracy of colour-based skin segmentation methods are significantly affected by different issues such as presence of skin-like colours in scene background and uncontrolled illumination conditions. Standard skin probability map (SPM) can not perfectly discriminate skin and non-skin regions in these conditions. To overcome limitations of SPM, a new probability map termed as discriminative space map (DSM) is proposed by extracting most discriminative features between skin and non-skin regions. A novel adaptive discriminative analysis (ADA) is proposed to extract most discriminant features between skin and non-skin regions from an image itself in an unsupervised manner. Subsequently, a dynamic region growing (DRG) method is employed to allow skin regions to grow dynamically. The DRG controls false detection by restricting the region growing process. Experimental results for standard databases show that the proposed method can efficiently segment out skin pixels in presence of skin-like background colours and uncontrolled illumination conditions.

2.1 Introduction

In-spite of having several advantages of colour-based skin segmentation, accurate segmentation of hand or any other body parts is still a big challenge. The accuracy of colour-based skin detection methods are severely affected by the presence of skin-like colours in the background. Majority of skin detection algorithms use skin colour as a primary feature. However, use of some other features like texture information [103] or depth information [104, 105] along with colour feature generally improve segmentation accuracy. Zhiwei *et al.* [83] used Gabor wavelet-based texture filter, whereas Dumitrescu and Dumitrache [106] used Gray Level Co-occurrence Matrix (GLCM) texture features for skin detection. In all these methods, it is assumed that skin regions are smoother than background regions. But, the background may have similar textural smoothness as that of skin regions. In these specific cases, chromatic parameters, such as hue, saturation may be the discriminative features between skin and non-skin regions. Kawulok *et al.* [9] used linear discriminative analysis (LDA) to derive most discriminative features between skin and non-skin regions. In this method, LDA projection matrix is derived using colour and local texture features from a set of labelled images. The LDA projection matrix depends on training data. Therefore, LDA gives a projection matrix which ensures the best possible inter-class discrimination. However, natural images do not have any fixed textural or chromatic pattern. Therefore, it is highly likely that the test image characteristics may be significantly different from the training images. In these specific cases, derived discriminative feature may not be the most discriminative for the test images. So, an image specific discriminative feature extraction method is essential, so that the image specific features can be extracted for discriminant analysis for skin segmentation. Kawulok *et. al.* [15] also proposed an adaptive scheme to select seeds to further improve the detection results. Hettiarachchi and Peters [16] showed that the accuracy of skin detection can be significanly improved using multi-manifold learning. However, multi-manifold learning-based method uses some pre-selected true skin and skin-like non-skin samples for training. But, discrimination between true skin and skin-like non-skin pixels entirely depends on scene characteristics. The methods proposed in [10, 21, 23] extract facial pixel to generate locally adapted skin model. However, a face must be detected in an image in order to apply these methods. Most of these methods follow unsupervised or semi-supervised learning approaches. However, recent trend in computer vision is Convolutional Neural Network (CNN)-based architectures for pixel-wise prediction, such as semantic segmentation using

fully convolutional networks (FCN) [107]. However, performance of CNNs in complex computer vision problems, such as pixel-wise classification is limited. The convolutional filters with large receptive fields in traditional CNNs result in coarse outputs for pixel-wise classification [107]. The result becomes coarser due to the presence of maxpooling layers in CNNs [108]. This results in smooth boundaries and blob-like structures in segmented images. Also, lack of smoothness constraint in CNNs may result in spurious regions with poor object outlines in the segmented images [109–111]. Zheng et al. [112] showed that incorporation of conditional random fields (CRFs) at the final layer of CNN can refine coarse predictions. However, skin detection is a very subjective problem which depends on image characteristics. A colour (e.g, brown) may be the skin colour for African people or it may be a background colour for Caucasian people. Thus, lack of image specific adaptivity in existing semantic segmentation algorithms may limit their applicability in skin detection.

As mentioned by Kawulok [9], LDA can be used to extract most discriminative features for skin and non-skin pixels. However, LDA needs a set of labelled training dataset. In skin colour segmentation problem, some of the training samples belonging to the non-skin regions, like background regions, look very similar to the actual skin regions, and vice-versa. As the training samples for two classes (skin and non-skin) are obtained from an image itself, and so, a few samples corresponding to background and other regions take the identity of samples of the actual skin colour regions, and vice-versa. So, there will be marginal inter-class overlapping of samples extracted from an unknown image. Therefore, a feature derived using LDA may not be most discriminative for an unknown image. To address this issue, a novel adaptive discriminative analysis (ADA) for skin segmentation is proposed. It is "adaptive" in a sense that the ADA adapts to inter-class similarity or overlapping samples. Hence, the extracted feature would be image specific, not generic for all the images. In our approach, two sets of pixel locations S_1 and S_2 are used for deriving ADA projection vector. The set S_1 contains pixel locations that are very likely to be skin, which implies that some of them may look like skin, but they actually belong to non-skin regions. The other set S_2 contains pixel locations that are mostly non-skin, but a few of them belonging to actual skin regions. Therefore, there is no labelled training dataset. The proposed ADA assigns class-memberships to each of the training samples in S_1 and S_2. Subsequently, a discriminative feature space is extracted from the given image itself with the help of the proposed ADA. Finally, a discriminative space map

(DSM) is derived by using the discriminative feature space. In contrary to standard SPM, the proposed DSM provides better discrimination between skin and non-skin regions of an image.

The skin regions can be detected either by directly thresholding the skin map or DSM [13]. However, direct thresholding of DSM may increase false detection rates when skin colour is very much similar to background scene colour. Seeded region growing (SRG) of skin regions is an alternative to direct thresolding of SPMs [113]. The use of SRG algorithm for image segmentation has been well investigated in recent years. Adams and Bischof [114] first proposed the concept of SRG. The major limitation of SRG algorithm is the selection of an appropriate threshold to judge a similarity condition, which limits its applicability in skin segmentation. If a very high threshold is selected, then many true skin pixels could be rejected. On the other hand, false detection will be more for a small threshold. Propagation-based region growing method as proposed by Kawulok [14] approximates an image as a graph with the pixels as nodes within it. Kawulok's method is based on Dijkstra's minimum path-cost algorithm [88]. The region growing method proposed by Kawulok (2013) is not adaptive. The entire region growing process depends on manually selected parameters. The region growing should be dependent on scene characteristics of an image. Region growing should be more if the background colour similarity with the skin colour is less, and vice-versa. To handle these issues, we proposed a *dynamic region growing (DRG)* method by employing only few manually selected parameters. The proposed method is discussed in detail in the Section to follow.

2.2 Proposed Method

Figure 2.1[1] shows the block diagram of our proposed skin detection method. In the first step of our proposed method, a discriminative space map (DSM) is derived to enhance the discrimination between skin and non-skin pixels, and derivation of DSM only requires a discriminative analysis. In our method, a novel adaptive discriminative analysis (ADA) method is proposed. ADA requires two sets of pixel locations. The first set contains most of the skin regions, and it may contain a few ambiguous non-skin regions. Similarly, the second set contains discernible non skin regions along with a few skin like coloured regions. The projection vector is derived on the basis of these two indistinct training sets. A generalized training set can not be extracted

[1]This work is under major revision in *Journal of Visual Communication and Image Representation*, Elsevier (Revised paper submitted).

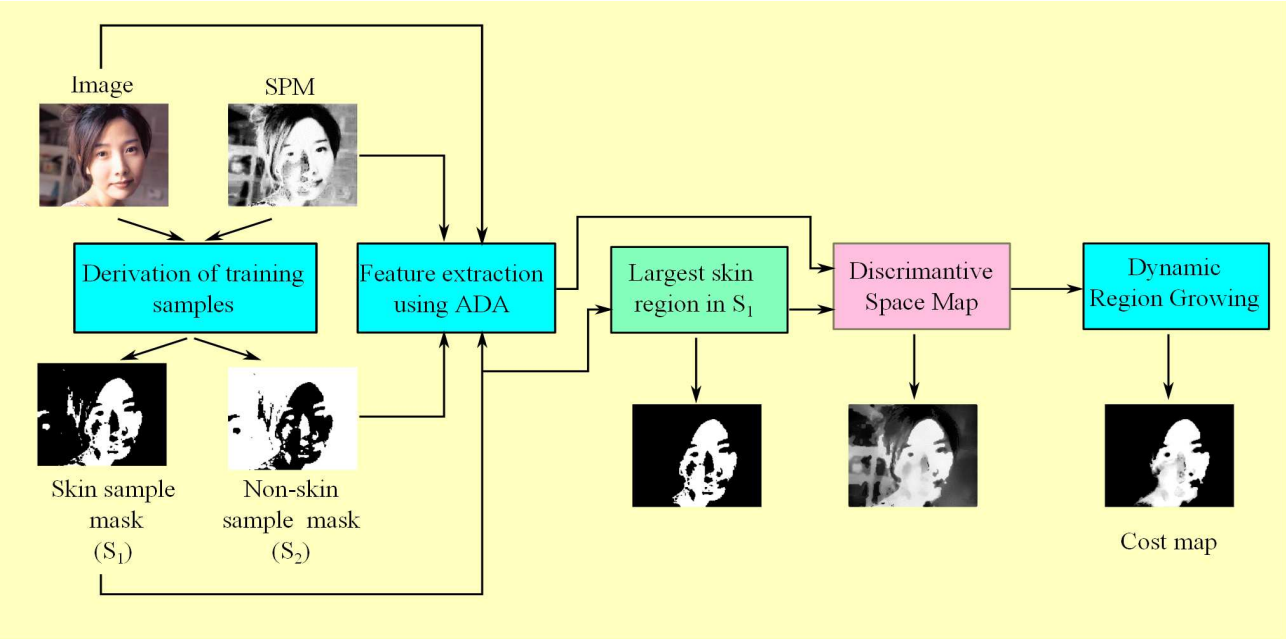


Figure 2.1: Block diagram of the proposed skin detection algorithm.

from different images due to skin-background colour similarity and non-uniform illumination, and so, our method relies on extracting image specific training samples for classification. For this, skin and non-skin sample masks are derived from a given image. These masks give skin and non-skin sample sets for the proposed ADA. So, ADA can extract a discriminative feature between skin and non-skin regions. A DSM is obtained by using the discriminative feature and the largest region in the skin sample mask. Subsequently, a dynamic region growing (DRG)[2] method using DSM is employed, which gives a cost map. Finally, the skin mask is obtained by thresholding of the estimated cost map.

2.2.1 Discriminative space map

In order to increase discrimination between skin and non-skin regions, a DSM can be derived by using SPM and other image features. The novelty of our method is the derivation of image specific discriminative features using ADA. The proposed ADA is trained with a set of highly likely skin and non-skin samples derived from the input image itself. Let, S_1 is a set of pixel locations which are mostly to be skin, and S_2 is a set of pixel locations which are mostly to be non-skin. It is observed that the skin regions possess higher SPM values as compered to the

[2]This work has been published in *Twenty Second National Conference on Communication (NCC), Guwahati, 2016.* (Refer item 2 in Page 117 for details)

non-skin regions. So, thresholding the skin probability map with a high threshold value (θ_{High}) can provide the required set S_1 [14]. However, the non-skin regions may have skin-like colours in some of the instances. In such cases, non-skin pixel locations having high SPM values could be wrongly included in S_1. In our previous work [103], it is shown that the use of texture information along with *Possibilistic Fuzzy C-Means* (PFCM) clustering [115] can improve the segmentation results. In order to reduce falsely included non-skin pixel locations into S_1, PFCM clustering is employed in our method as follows:

- Divide the SPM into three set of regions as: a) regions with highly likely skin-belonging; b) ambiguous skin regions; and c) regions with highly-likely non-skin belonging. In order to obtain the three set of regions, we apply Otsu's multi-thresholding algorithm [116] on SPM. It gives two threshold values $[\theta_{Low}, \theta_{High}]$.

- Let S_{old} is the initial set of skin-like pixel locations x in an image such that

$$x \in \{S_{old} : SPM(x) > \theta_{High}\}, \tag{2.1}$$

  In some of the cases, S_{old} may contain skin-colour like non-skin locations due to skin-background colour similarity. Therefore, the PFCM clustering algorithm is applied on the set S_{old} to group and discard non-skin pixel locations within it.

- In our method, two types of features are selected for PFCM algorithm, namely – colour features and texture features. The colour features include luminance values (Y), chrominance values in YCbCr domain (Cb, Cr), and hue (H). The texture features include entropy, standard deviation, and maximum-minimum difference. The texture features are extracted from SPM [9] by using the kernels of sizes 3×3.

- The centroid of a cluster derived from PFCM clustering is expressed as:

$$\mathbf{V}_i = \frac{\sum_{j=1}^{N} \left(a u_{ij}^m + b t_{ij}^\eta\right) \mathbf{f}(x_j)}{\sum_{j=1}^{N} \left(a u_{ij}^m + b t_{ij}^\eta\right)}, i = 1, 2. \tag{2.2}$$

where,

$$\begin{aligned} u_{i,j} &= \text{membership value of the } j^{th} \text{ pixel in the } i^{th} \text{ cluster} \\ t_{i,j} &= \text{typicality value of the } j^{th} \text{ pixel in the } i^{th} \text{ cluster} \\ \mathbf{f}(x_j) &= \text{feature vector at } j^{th} \text{ location } x_j, \\ N &= \text{Total number of input sample } \mathbf{f}(x_j), \\ a, b &= \text{positive constants}, \\ m, \eta &= \text{weighting exponent} \end{aligned}$$

Let, the i^{th} cluster is denoted by Ω_i. Therefore, final clustering of the input samples is done as follows:

$$x_j \in \Omega_i : i = \arg\min_{\forall k} \|\mathbf{f}(x_j) - \mathbf{V}_k\|, \ k = 1, 2. \tag{2.3}$$

Finally two binary images are obtained from these two clusters such that,

$$BW_i(x_j) = \begin{cases} 1 \text{ if } x_j \in \Omega_i \\ 0 \text{ otherwise} \end{cases} \tag{2.4}$$

- It is expected that the cluster Ω_i having most of the true skin pixels should be statistically more nearer to the global skin pixel distribution model [117]. The global skin distribution model is obtained from the training skin samples for derivation of SPM. The global skin distribution model can be modelled using GMM as:

$$P(\mathbf{X}|Skin) = \sum_{k=1}^{K_{Skin}} \omega_k^{Skin} \cdot G\left(\mu_k^{Skin}, \mathbf{\Sigma}_k^{Skin}\right) \quad , \tag{2.5}$$

where, ω_k^{Skin} is the prior probability weight, K_{Skin} is the number of Gaussians used for GMM, μ_k^{Skin} is the mean, and Σ_k^{Skin} covariance matrix of the k^{th} Gaussian $G\left(\mu_k^{Skin}, \mathbf{\Sigma}_k^{Skin}\right)$. Similarly, the distribution of pixels in RGB domain for the i^{th} binary image BW_i can be represented as $G\left(\mu_i^{BW}, \mathbf{\Sigma}_i^{BW}\right)$ by using a single Gaussian function. We use the notation $G_i \equiv G\left(\mu_i, \mathbf{\Sigma}_i\right)$ in the rest of this chapter for simplicity. It is highly likely that the number of skin pixels are significantly large as compared to non-skin pixels in S_{old} in some of the cases. In those situations, the binary image with mostly non-skin pixels indicate

higher percentage of pixels belonging to the edges. In some of the cases, skin-like colours may cover majority of background regions. It is observed that these background regions also include a significant amount of image boundary pixels. Hence, the binary image containing true-skin pixels is selected based on following decision rule:

$$BW_{final} = \begin{cases} BW_1 \text{ if } \left\{ N_2^b \geqslant \dfrac{w+h}{2} \right\} \cup \left\{ \left(e_1 \cdot d_1^{Skin} \leqslant e_2 \cdot d_2^{Skin} \right) \cap \left(N_i^b < \dfrac{w+h}{2}, \forall i \right) \right\} \\ BW_2 \text{ otherwise} \end{cases} \tag{2.6}$$

where,

$$\begin{aligned} d_i^{Skin} =& \frac{1}{K_{Skin}} \sum_{k=1}^{K_{Skin}} D\left(G_k^{Skin}, G_i^{BW}\right), \\ D\left(G_k^{Skin}, G_i^{BW}\right) =& \text{Bhattacharya distance between } G_k^{Skin} \text{ and } G_i^{BW} \\ =& \frac{1}{8}\left(\mu_k^{Skin} - \mu_i^{BW}\right)^T \left[\frac{\Sigma_k^{Skin} + \Sigma_i^{BW}}{2}\right]^{-1} \left(\mu_k^{Skin} - \mu_i^{BW}\right) \\ &+ \frac{1}{2} \ln \frac{\left|\left(\Sigma_k^{Skin} + \Sigma_i^{BW}\right)/2\right|}{\sqrt{\left|\Sigma_k^{Skin}\right| \cdot \left|\Sigma_i^{BW}\right|}}, \\ e_i =& \% \text{ of edge pixels in } BW_i \text{ for } i = 1, 2. \\ N_i^b =& \text{number of image boundary pixel locations in } BW_i \text{ for } i = 1, 2. \\ w \times h =& \text{image dimension} \end{aligned} \tag{2.7}$$

- Hence, S_1 is obtained as:

$$S_1 = \{x : BW_{final}(x) = 1\} \tag{2.8}$$

Figure 2.2 shows the block diagram of proposed method for obtaining skin sample set S_1 for ADA. Again, the set of non-skin or skin-like pixel locations S_2 can be obtained as follows:

$$S_2 = \{x : BW_{final}(x) = 0\} \tag{2.9}$$

In order to obtain a discriminitive feature space, we need to obtain a suitable projection vector to increase inter-class discrimination. In-spite of PFCM-based refinement[3], there is

[3]This work has been already published in *IEEE Recent Advances in Intelligent Computational Systems (RAICS), Trivandrum, 2015* (Refer item 4 in Page 117 for details)

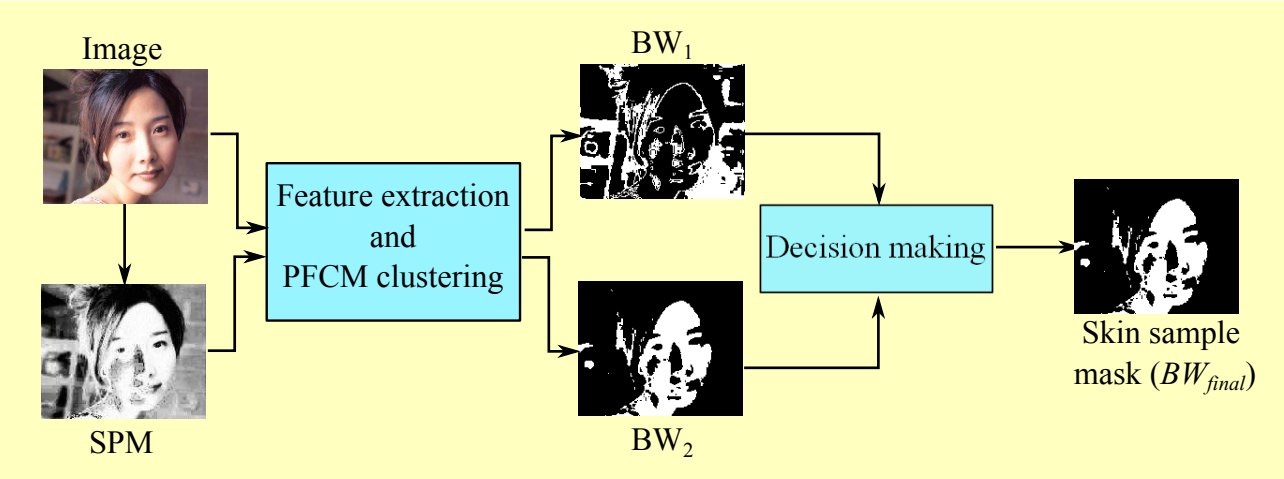


Figure 2.2: Proposed method for obtaining the set S_1

a possibility that some of the true non-skin pixels are included in S_1. On the otherhand, some of the true skin pixels may be included in S_2 due to their higher similarity to non-skin regions. These wrongly included pixels act as outliers in their respective sets. So, a membership assignment method is proposed for each of the samples in their respective clusters. The membership values are calculated in such a manner that the membership of the outliers in their respective clusters become very small. In our approach, membership value of a sample is calculated using its mean intra-cluster distance and mean inter-cluster distance. The proposed ADA algorithm is described as follows:

Consider the pixel locations $x_{1,k} \in S_1$ and the pixel locations $x_{2,l} \in S_2$, where $k = 1, .., N_1$, and $l = 1, .., N_2$. Here, N_1 and N_2 are the number of samples in S_1 and S_2, respectively. Therefore, *membership* values for the pixels are obtained by:

$$\begin{aligned} \lambda_{1,k} &= \frac{N_1 - 1}{N_2} \times \frac{\sum_{l=1}^{N_2} d\left(x_{1,k}, x_{2,l}\right)}{\sum_{k'=1,k' \neq k}^{N_1} d\left(x_{1,k}, x_{1,k'}\right)} \\ \lambda_{2,l} &= \frac{N_2 - 1}{N_1} \times \frac{\sum_{k=1}^{N_1} d\left(x_{2,l}, x_{1,k}\right)}{\sum_{l'=1,l' \neq l}^{N_2} d\left(x_{2,l}, x_{2,l'}\right)} \end{aligned} \tag{2.10}$$

where,

$$\begin{aligned} d\left(x_{i,k}, x_{j,l}\right) =& [\mathbf{f}(x_{i,k}) - \mathbf{f}(x_{j,l})]^T \boldsymbol{\Sigma}_j^{-1} \left[\mathbf{f}(x_{i,k}) - \mathbf{f}(x_{j,l})\right], \text{ for } i,j = 1,2; \forall k,l \\ \boldsymbol{\Sigma}_j =& \text{Covaraince matrix of samples in } S_j,\ j = 1,2. \end{aligned} \tag{2.11}$$

The mean vectors of two classes are given by:

$$\boldsymbol{\mu}_1 = \frac{\sum_{k=1}^{N_1} \lambda_{1,k} \cdot \mathbf{f}\left(x_{1,k}\right)}{\sum_{k=1}^{N_1} \lambda_{1,k}} \quad \text{and} \quad \boldsymbol{\mu}_2 = \frac{\sum_{l=1}^{N_2} \lambda_{2,l} \cdot \mathbf{f}\left(x_{2,l}\right)}{\sum_{l=1}^{N_2} \lambda_{2,l}} \tag{2.12}$$

The *within-class* scatter matrix is obtained by:

$$\mathbf{S_w} = \frac{\sum_{k=1}^{N_1} \lambda_{1,k} \cdot \left(\mathbf{f}\left(x_{1,k}\right) - \boldsymbol{\mu}_1\right) \cdot \left(\mathbf{f}\left(x_{1,k}\right) - \boldsymbol{\mu}_1\right)'}{\sum_{k=1}^{N_1} \lambda_{1,k}} + \frac{\sum_{l=1}^{N_2} \lambda_{2,l} \cdot \left(\mathbf{f}\left(x_{2,l}\right) - \boldsymbol{\mu}_2\right) \cdot \left(\mathbf{f}\left(x_{2,l}\right) - \boldsymbol{\mu}_2\right)'}{\sum_{l=1}^{N_2} \lambda_{2,l}} \tag{2.13}$$

where, $\mathbf{f}\left(\bullet\right)$ is the feature vector. The feature vector for ADA comprises of texture features (entropy, standard deviation, maximum-minimum difference) and colour features (SPM, Cb-Cr components and H values). The texture features are derived from SPM by using the kernels of sizes 3×3, 7×7, and 11×11. Therefore, the *between-class* scatter matrix is given by:

$$\mathbf{S_b} = \left(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}\right)\left(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}\right)' \tag{2.14}$$

The optimal ADA projection vector $\mathbf{Z}$ for the input samples is obtained as the eigenvector of the matrix $\mathbf{S}_b^{-1}\mathbf{S}_w$ corresponding to its highest eigenvalue. It is observed that the largest region (LR) in BW_{final} corresponds to true skin regions in most of the instances. Hence, the discriminative space map of an image is given by:

$$DSM(x) = \left|\mathbf{Z}' \cdot \left(\mathbf{f}(x) - \boldsymbol{\mu}_{LR}\right)\right| \tag{2.15}$$

where, $\boldsymbol{\mu}_{LR}$ is the centroid of the largest region in BW_{final} in feature space.

To show the effectiveness of our proposed method, discriminative maps obtained by Kawulok's method [9], and the proposed DSM are compared as shown in Figure 2.3. From the results of this analysis, it is clearly observed that Kawulok's discriminative map is not adaptive to the changes in the scene characteristics. On the other hand, there is no need of labelled training

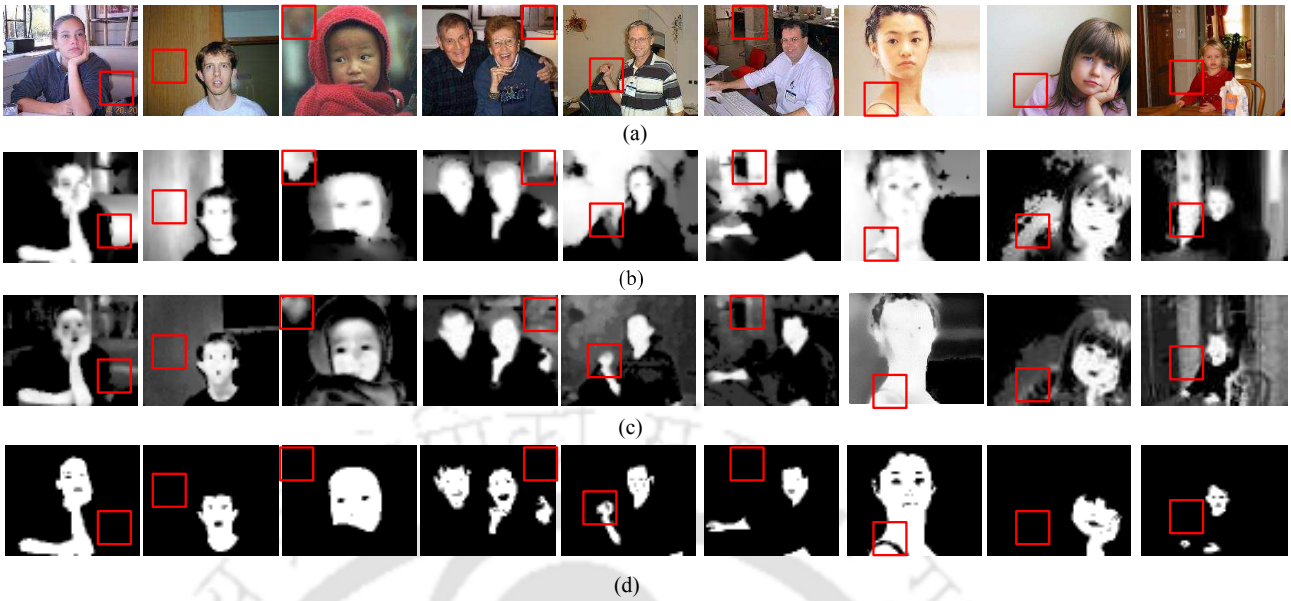


Figure 2.3: Experimental results showing discrimination between skin and non-skin pixels: a) Original image, b) DSPF [9], c) Proposed Discriminative Space Map (DSM), and d) Ground-truth Results. The red coloured boxes show the specific regions where our proposed method gives comparatively better segmentation results.

samples to obtain our proposed DSM. In our method, the test image itself is used to derive the projection matrix, and hence, discrimination is more and it is adaptive to scene changes.

The final skin mask for a given image can be easily obtained by direct thresholding of the DSM with a suitable threshold. However, direct thresholding may consider some non-skin regions with comparatively higher DSM values as skin. This increases the chance of false positive error. Therefore, a seeded dynamic region growing (DRG) method with dynamically controlled region expansion is used. The DRG significantly reduces false positive errors. The detailed discussion on DRG is given in the next section.

2.2.2 Dynamic region growing

In this work, a novel dynamic region growing (DRG) method is proposed, which shows improvement over the existing method [118]. In the proposed DSM, a lower value of $DSM(x)$ corresponds to a higher probability of obtaining skin coloured pixels. For region growing, a set of seed pixels are needed to be extracted from the DSM. To obtain the seed pixels, the DSM

needs to be inverted. The inverted DSM is obtained as:

$$DSM(x) \Leftarrow \frac{DSM_{\max} - DSM(x)}{DSM_{\max}} \tag{2.16}$$

where, $DSM_{\max}$ is the DSM maximum. In order to get more discrimination between skin and skin-like non-skin regions, locations $\{x : SPM(x) < 0.1\}$ are discarded during calculation of $DSM_{\max}$ and the inverted DSM.

The seed pixels for DRG is obtained by thresholding the DSM with a threshold P_s [9, 23]. As described by Kawulok [14], the local cost to reach a non-visited pixel x depends on two factors: a) cost incurred in the colour domain, and b) cost incurred in the DSM domain. So, the local cost at pixel location x is given by:

$$C_l(x) = \begin{cases} C_I(x) \cdot C_p(x) \text{ for } DSM(x) \geq P_{th} \\ \infty \text{ otherwise} \end{cases} \tag{2.17}$$

where, P_{th} is the limiting value of the DSM. In equation (2.17), $C_I(x)$ is the local cost in colour space and $C_p(x)$ is the local cost in DSM space. It is observed that the skin regions are more dicriminative in chromatic domain than in luminance domain. So, Y, Cb, and Cr components of pixels are chosen for $C_I(x)$ calculation. Therefore, the local cost in colour domain at the i^{th} pixel location q_i in the path *i.e.* $C_I(q_i)$ is given by:

$$C_I(q_i) = \frac{1}{3}\left[|Y(q_i) - Y(q_{i-1})| + |Cb(q_i) - Cb(q_{i-1})| + |Cr(q_i) - Cr(q_{i-1})|\right], \tag{2.18}$$

The local cost in DSM domain should be less sensitive to smaller values of $P(q_i)$. On the other hand, the cost function $C_p(x)$ should increase the cost more aggressively for higher values of $P(q_i)$. Therefore, the local cost in the DSM domain is expressed as:

$$C_p(q_i) = \left(1 + \Delta_{q_i} + \Delta_{q_i}^2\right) e^{\Delta_{q_i}^2} \quad . \tag{2.19}$$

where, $\Delta_{q_i} = P_s - P(q_i)$.

The parameter P_{th} controls growth of the regions. Therefore, the detection error is significantly dependent on the value of P_{th}. The value of P_{th} can be selected manually [14] or adaptively [118]. Adaptive selection of P_{th} can improve detection result significantly [118].

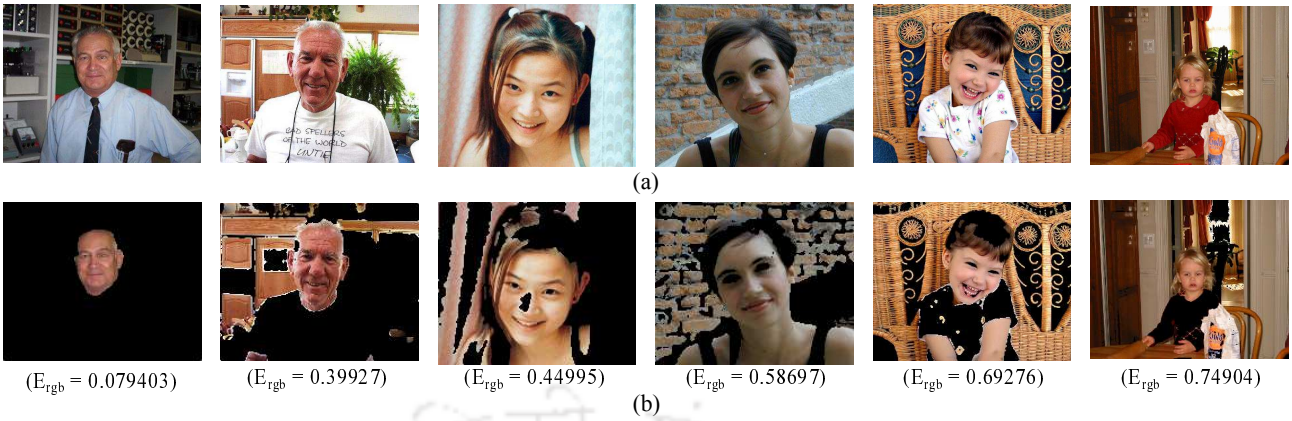


Figure 2.4: Illustration of skin chroma entropy as a measure of possible false acceptance error: a) Original image, and b) image pixels with $DSM(y) \geqslant P_l$ and associated E_s values.

Hence, a method is proposed to automatically determine P_{th}. Our proposed method to determine P_s and P_{th} are discussed in the next paragraph.

First, the DSM is divided into three parts using Otsu's multi-thresholding algorithm [116]. Subsequently, we obtained two thresholds (P_h, P_l). The regions in DSM with $DSM(x) > P_h$ are highly likely to be skin only if the colour similarity between skin and the background is not so prominent. On the other hand, the regions in DSM with $DSM(x) < P_l$ are highly likely to be non-skin. Hence, the value of P_{th} should be in the range of $[P_l, P_h]$. However in uncontrolled environments, background regions could be chromatically and texturally similar, and for such cases, DSM even can not perfectly discriminate skin and non-skin regions. Therefore, P_{th} should be in the range of $[P_l, P_s]$. The thresholds P_s and P_{th} play a significant role in regulating detection errors. If both of them are very low, the false acceptance increases in presence of skin like background colours. On the otherhand, if both of them are too high, false rejection could be high if skin regions are not uniformly illuminated. A parameter termed as skin-chroma entropy (E_c)[4] for a given colour space c, is employed to characterize the thresholds (see Figure 2.4). The skin-chroma entropy is calculated for RGB colour space, and it is denoted as E_s. The parameter E_s can be obtained as:

$$E_s = -\frac{\sum\limits_{\forall y \in \mathbf{Y}} n_{\mathbf{X}(y)} \log\left(\frac{n_{\mathbf{X}(y)}}{N_{\mathbf{Y}}}\right)}{N_{\mathbf{Y}} \cdot \log(255^3)} \tag{2.20}$$

[4]This work is published in *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2016.* (Refer item 3 in Page 117 for details)

where,

$$y \in \{\mathbf{Y} : DSM(y) \geqslant P_l\} \tag{2.21}$$

Here, $n_{\mathbf{X}(y)}$ is the number of count of colour vector $\mathbf{X}(\mathbf{y})$ and $N_{\mathbf{Y}}$ is the total number of pixel locations in $\mathbf{Y}$. We express the threshold P_s as a function of E_s as:

$$P_s = P_h + E_s \cdot (P_{s,\max} - P_h)\, e^{1-E_s} \tag{2.22}$$

where, $P_{s,\max}$ is the maximum allowable value of P_s.

Subsequently, the parameter P_{th} can also be expressed as a function of E_s as:

$$P_{th} = P_0 e^{\beta(1-E_s)} + P_l \tag{2.23}$$

where,

$$P_0 = (P_s - P_l) \cdot E_s \tag{2.24}$$

The parameter β controls the rate of increment of P_{th} with respect to E_s. From eq. (2.22) and (2.23) it is clear that,

$$\left.\begin{aligned} P_s &\approx P_h \\ P_{th} &\approx P_l \end{aligned}\right\} \text{ for } E_s \approx 0 \quad , \tag{2.25}$$

and

$$\left.\begin{aligned} P_s &\approx P_{s,\max} \\ P_{th} &\approx P_{s,\max} \end{aligned}\right\} \text{ for } E_s \approx 1 \tag{2.26}$$

Finally, the path costs for all the pixel locations x are calculated and expressed as a cost map for the image as,

$$\begin{aligned} C(x) =& \sum_{i=1}^{n} C_l\left(q_{i-1} \rightarrow q_i\right) \\ =& C_l(x) + C\left(q_{n-1}\right) \end{aligned} \tag{2.27}$$

The final skin mask is obtained by thresholding the cost map $C(x)$ with a threshold θ_{th}.

2.3 Experimental Results

2.3.1 Experimental setup

For validation of our proposed method, two datasets namely HGR [9,14,67], and ECU [65] are used. The HGR dataset contains 899 images taken under controlled illumination conditions. The ECU dataset has two parts, each containing 2000 images. These images are captured in uncontrolled natural environments. Out of two parts of ECU dataset, one part (2000 images) is used for GMM training to get the SPM. The HGR dataset and the remaining part of ECU dataset are used for experimental validation of the proposed method. For the PFCM, we choose the parameters, such as a, b, m, η according to [115] as $a = 1$, $b = 1$, $m = 2$, and $\eta = 2$. We select $P_{s,max} = 0.9$ judiciously for both of the datasets. Rest of the parameters except β are generic, and they are derived along the course of the process. In practical scenario, region growing may not stop for smooth regions with skin-like colour. Therefore, the maximum cost incurred should be bounded by some value τ. After a large number of experiments, the parameter τ is fixed at 40.

In this experimental validation process, four measures are selected for evaluation, namely - *detection accuracy (Acc.)*, *false positive error rate* (δ_{fp}), *false negative error rate* (δ_{fn}), *total detection error rate*(δ_t), where $\delta_t = \delta_{fp} + \delta_{fp}$. All the detection error values are obtained by thresholding the cost map with a threshold θ_{th}. Here, θ_{th} corresponds to minimum δ_t.

To derive the optimum value of β, detection errors are calculated for different values of β for both of the datasets. Figure 2.5 shows the variations of δ_t with β in different datasets. It is evident that at $\beta \simeq 0$ the error δ_t becomes minimum for both the datasets. This reduces the eq. (2.23) into,

$$P_{th} \simeq E_s \cdot (P_s - P_l) + P_l \tag{2.28}$$

2.3.2 Experimental validation

In our first analysis, performance of our proposed ADA is compared with the classical LDA. For this, DSMs are derived for images by replacing ADA with the traditional LDA in our proposed method. The comparative results are given in Table 2.1. From the results shown in Table 2.1, it is evident that ADA gives better inter-class discrimination than LDA for inter-class overlapping.

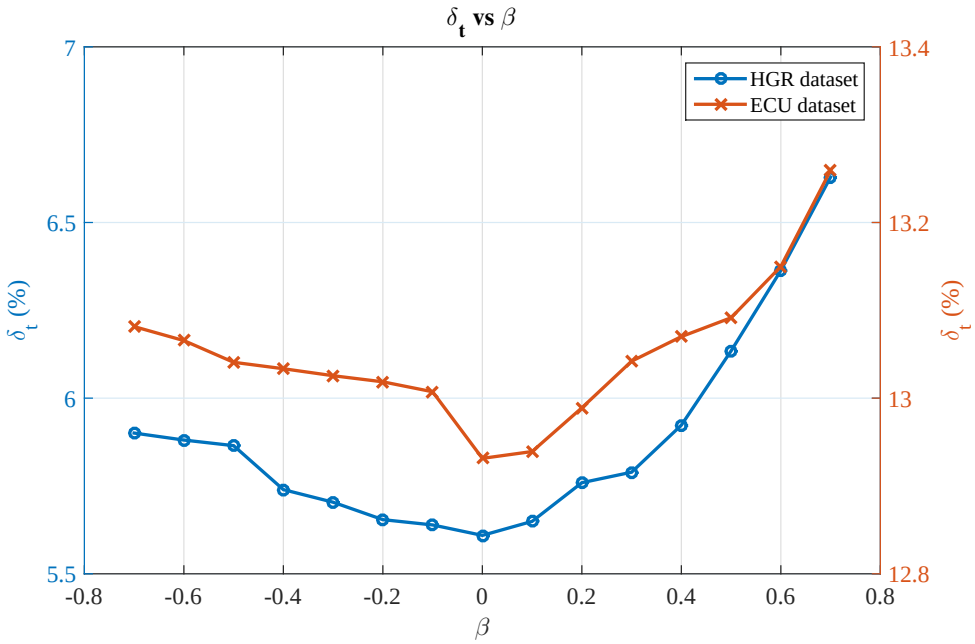


Figure 2.5: Variation of δ_t with β in different datasets

Table 2.1: Comparison between LDA and ADA in discriminative feature extraction

Method	HGR database				ECU database			
	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	Acc.	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	Acc.
SPM + direct threshold	5.47	5.15	10.62	0.9470	15.14	9.48	24.62	0.8803
DSM using LDA + direct threshold	1.78	5.15	6.93	0.9657	9.56	7.69	17.25	0.9051
DSM using ADA + direct threshold	1.21	4.53	5.74	0.9783	8.60	7.84	16.44	0.9128

In our second analysis, we investigated the overall improvement in skin detection due to the introduction of ADA. For that, detection errors are directly computed for the skin mask BW_{final} after PFCM, and compared with that produced from thresholded DSMs. The comparative analysis is given in Table 2.2 for different datasets. From the results given in Table 2.2, it is observed that the mask BW_{final} captures many true skin regions with very less false acceptance error. However, it fails to detect all the skin regions due to colour variations on account of non-uniform illumination as shown in Figure 2.6. This problem is somewhat resolved with the help of DSM, which relies on ADA. From Figure 2.6, it is visible that DSM can successfully identify more true skin regions as compared to BW_{final}.

The test results for some sample images taken from both HGR and ECU datasets are shown in Figure 2.7 and Figure 2.8. It is observed that the proposed algorithm is able to extract a skin

Figure 2.6: Intermediate test results for some sample images selected from HGR and ECU datasets.

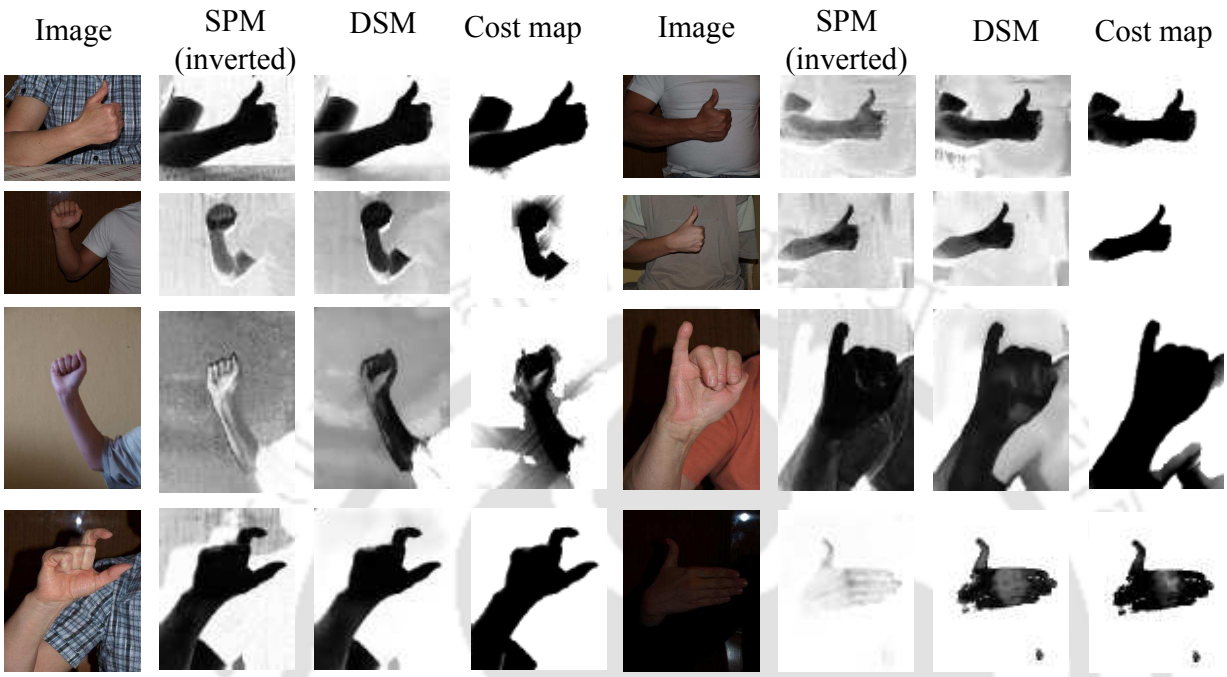


Figure 2.7: Test results for some sample images selected from HGR dataset.

Table 2.2: Comparison of skin detection between BW_{final} and thresholded DSM

Method	HGR database				ECU database			
	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	Acc.	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	Acc.
BW_{final}	0.02	46.55	46.57	0.9007	3.38	27.28	30.66	0.9158
DSM + direct threshold	1.21	4.53	5.74	0.9783	8.60	7.84	16.44	0.9128

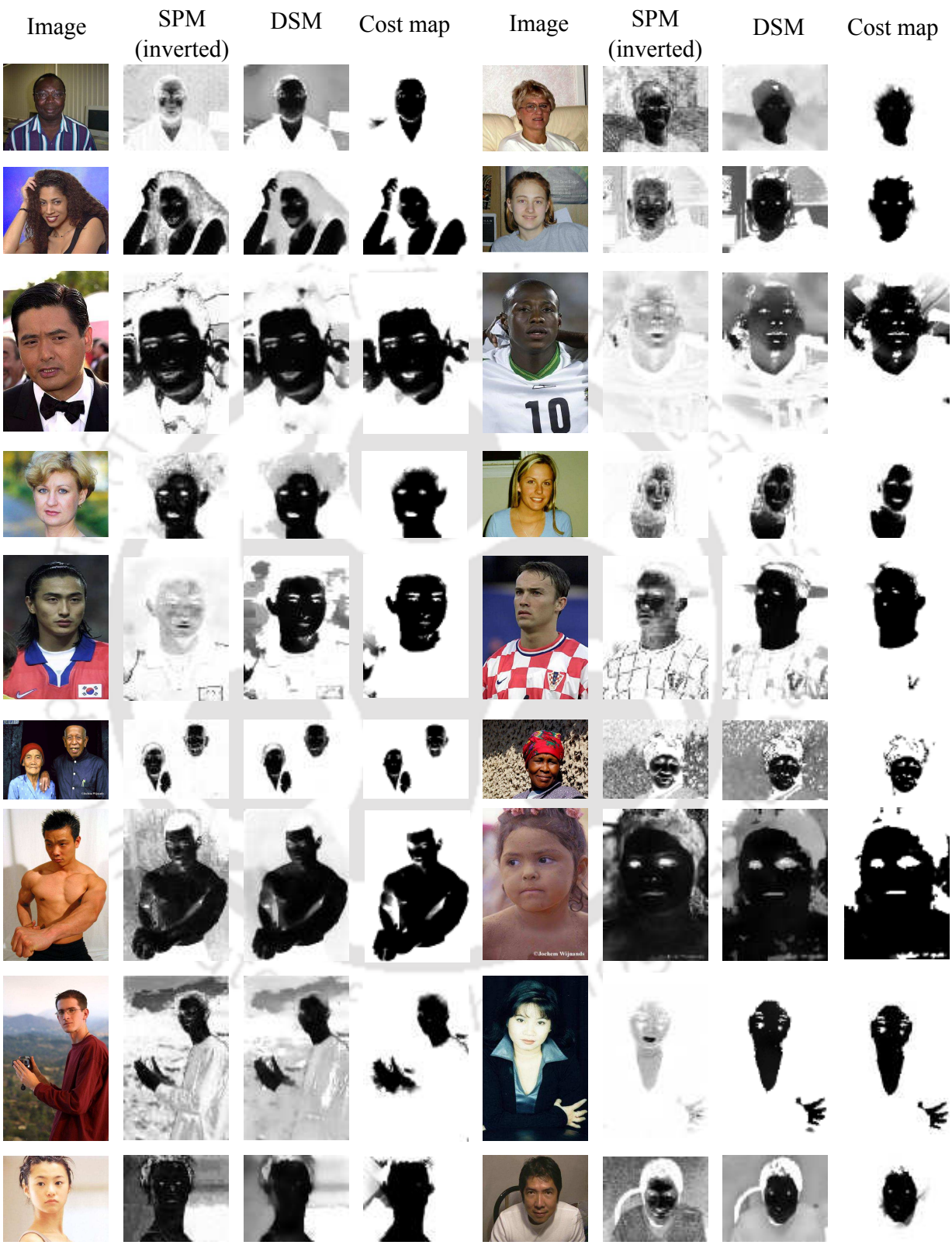


Figure 2.8: Test results for some sample images selected from ECU dataset.

map which is quite robust to different scenic changes in the images. Figure 2.8 also shows that the proposed skin map is not biased towards any racial skin tone. The accuracies of the methods which directly use SPM as a discriminative map for skin and non-skin pixels significantly depend on SPM's performance in detecting true skin regions. An SPM's discriminative ability depends on the quality of the training samples. Therefore, its performance degrades in poor illumination conditions or presence of skin-like coloured regions in the backgrounds. However, the proposed method adapts to local image characteristics, and it gives a more image specific discriminative map (DSM). From Figure 2.9, it is clear that the DSM is able to identify skin regions comparatively more accurately than the SPM under uneven illumination conditions.

Finally, the proposed segmentation method is compared with a number of well known state-of-the-art methods as given in Table 2.3. The performance of other methods are directly obtained from the results shown by Hettiarachchi and Peters [16] and Xu *et al.* [119]. It is evident from the results shown in Table 2.3 that the incorporation of DSM in our method significantly reduces overall detection error rates. The DSPF-based method outperforms Bayes method [13], FPSS [14], FNT [119] and SASC [22] for both HGR dataset and ECU dataset. In Kawlauk's method [9], discriminative features are extracted by LDA. However, the discriminative feature derived from LDA may not be the most discriminative for all the test images. This is due to the fact that the LDA tries to find out most discriminative features from the training sample sets, and the training sets of LDA are taken from the randomly selected skin and non-skin samples. If most of the training samples for two classes are well discriminated on account of texture features, then the projection vector will align in a direction for which texture discrimination is more. The same is also true for colour feature. Kawulok *et al.* also proposed a self-adaptive seed selection (SASS) region growing algorithm [15]. In this method, seed pixels derived from DSPF are used, and a local skin model is derived from these pixels. The SASS method performs better than DSPF for ECU dataset, whereas DSPF method performs better than SASS for HGR dataset. In the HGR dataset, there are many images where background and skin regions are chromatically very similar. In these cases, derivation of a reference pixel set by the expansion of seed regions may result in inclusion of some non-skin pixels into the reference pixel set. So, the chance of false positive error increases. In contrast to HGR dataset, the skin and the non-skin regions are chromatically less similar in most of the images of ECU dataset. In these cases, a local skin model derived from the reference pixels gives better skin detection

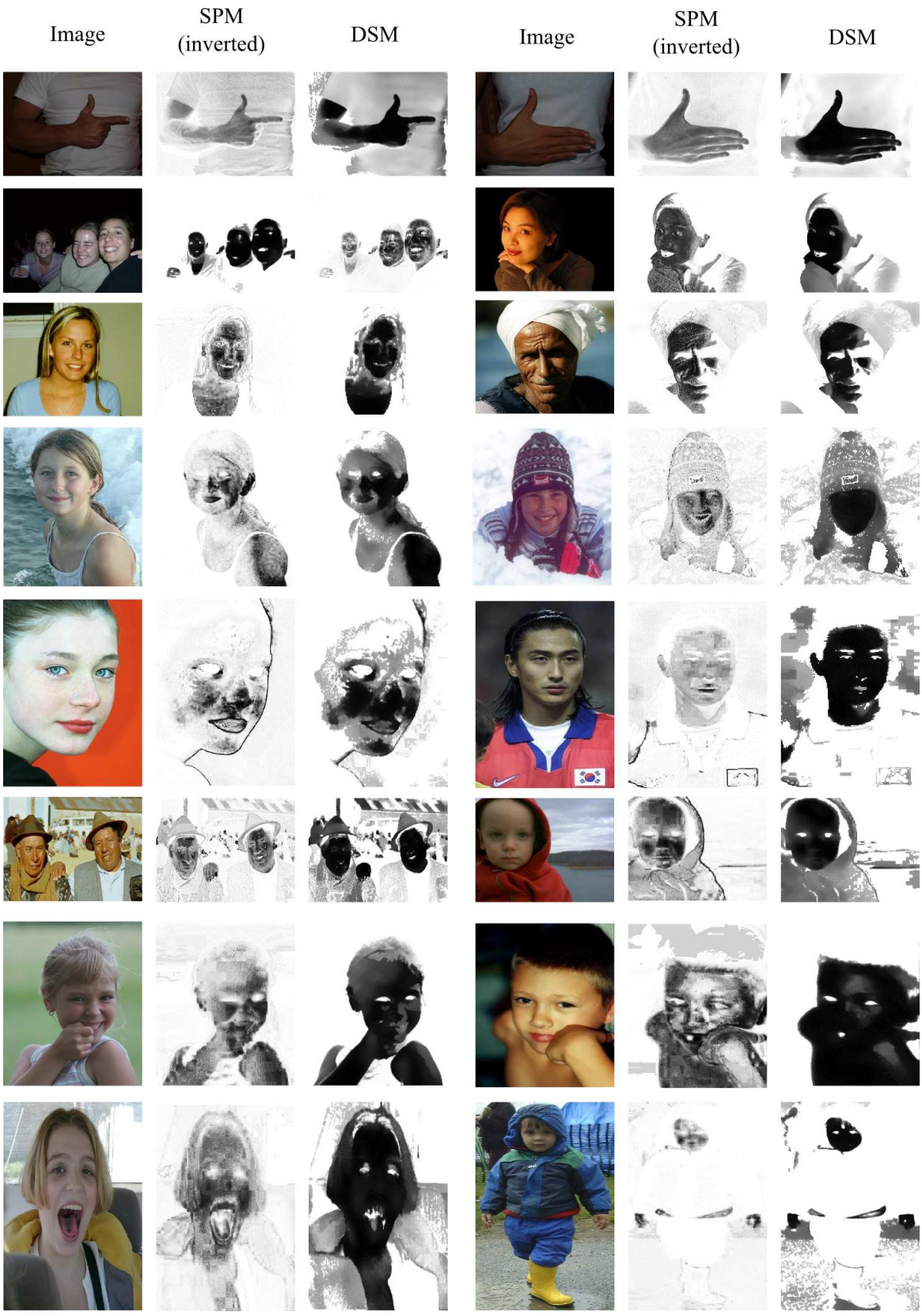


Figure 2.9: Test results for some sample images under varying illumination conditions

accuracy. The MMSC algorithm [16] outperforms all the previously mentioned methods for both the datasets. However, the MMSC-based method is not adaptive to the changes of scene characteristics. The ASSC algorithm [21] used face for localising seed points. Therefore, the ASSC method outperforms methods which do not use face detection for localising reference skin pixels. However, this method relies on standard SPM for region growing. Only seeds are extracted from a local skin model derived from facial pixels. The stacked auto encoder-based skin classification (SASC) method [22] uses a deep network comprising 3-stage autoencoder layer followed by a softmax layer. Each of the three autoencoders has 200 hidden neurons. The proposed DSM-based skin detection algorithm outperforms all the previously mentioned methods. To test the efficacy of DRG, cost maps are derived from SPM using the proposed DRG algorithm. The DRG algorithm outperforms FPSS algorithm for both the datasets. ROC plots for different methods for both HGR and ECU datasets are shown in Figure 2.10-2.11.

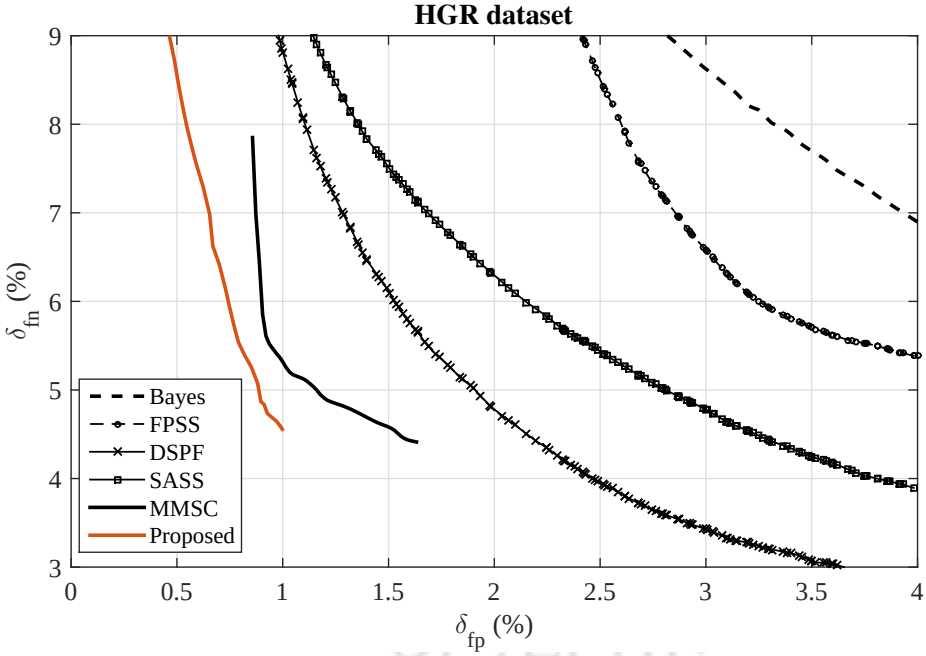


Figure 2.10: Comparative ROC curves HGR dataset

In general, the total area corresponding to the skin coloured regions is somewhat less than the non-skin regions. So, a skin detection method needs to produce less false acceptance error than false rejection error. It is evident from the ROC curves that the optimum operating region of the proposed method is located at a region in the plots for $\delta_{fp} \leq \delta_{fn}$. In Figure 2.12, test results of different methods are qualitatively compared with the results of our proposed method. For this, a set of sample images are selected from HGR and ECU dataset. From Figure 2.12, it is

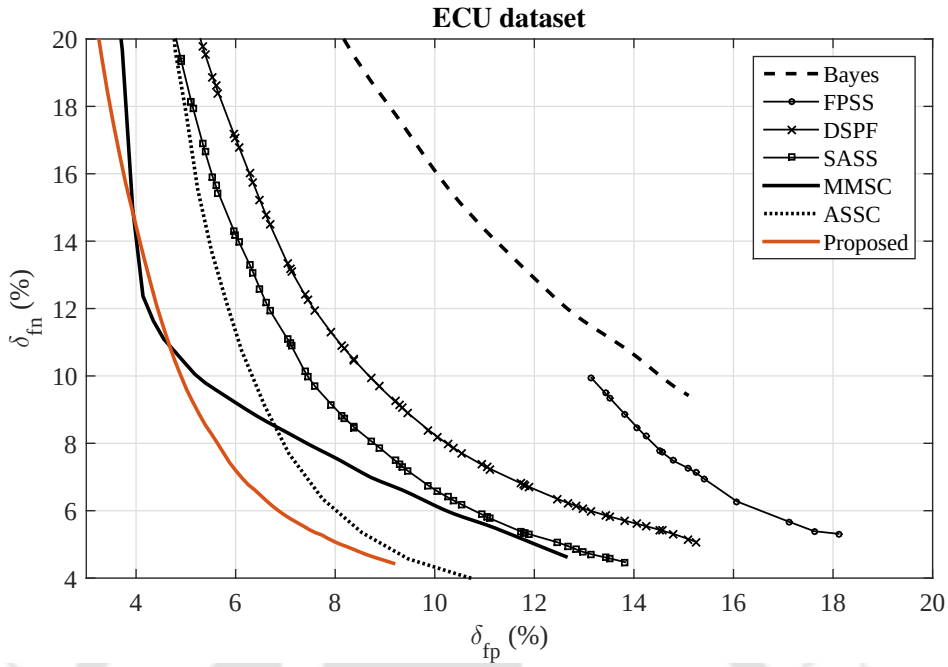


Figure 2.11: Comparative ROC curves ECU dataset

Table 2.3: Comparative analysis of different segmentation methods for HGR and ECU dataset

Method	HGR dataset				ECU dataset			
	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	**Acc.**	δ_{fp} (%)	δ_{fn} (%)	δ_t (%)	**Acc.**
Bayes classifier[1]	5.47	5.15	10.62	0.9470	15.14	9.48	24.62	0.8803
SASC[2]	4.00	20.00	24.00	0.9300	10.00	21.40	31.40	0.8800
FPSS[3]	3.46	5.72	9.18	0.9536	15.05	7.31	22.35	0.8924
FNT[4,†]	–	–	–	–	9.96	12.86	22.82	0.9053
DSPF[5]	2.24	**4.34**	6.58	0.9668	9.31	9.10	18.41	0.9080
SASS[6]	3.16	4.57	7.72	0.9611	9.53	7.16	16.68	0.9175
MMSC[7]	1.66	4.41	6.07	0.9692	**4.45**	11.43	15.74	0.9177
ASSC[8]	-	-	-	-	8.41	**5.88**	14.30	0.9184
SPM + **DRG**	2.71	5.37	8.03	0.9680	6.60	8.61	15.21	0.9291
DSM + DRG	**0.98**	4.63	**5.61**	**0.9818**	6.37	6.56	**12.93**	**0.9310**

[1]Bayes classifier [13]
[2]Stacked Autoencoders-based skin classification (SASC) [22]
[3]Fast proppagation-based skin segmentation (FPSS) [14]
[4]Flexible neural tree (FNT) [119]
[5]Discreminative Skin-Presence Feature (DSPF) [9]
[6]Self-adaptive seed selection (SASS) [15]
[7]Multi manifold-based skin classification (MMSC) [16]
[8]Adaptive seed-based skin classification (ASSC) [21]
[†]Result is unavailable for HGR dataset .

visible that the proposed skin detection algorithm is capable of segmenting out skin coloured regions more effectively than the state-of-the-art methods even if there is a skin-background

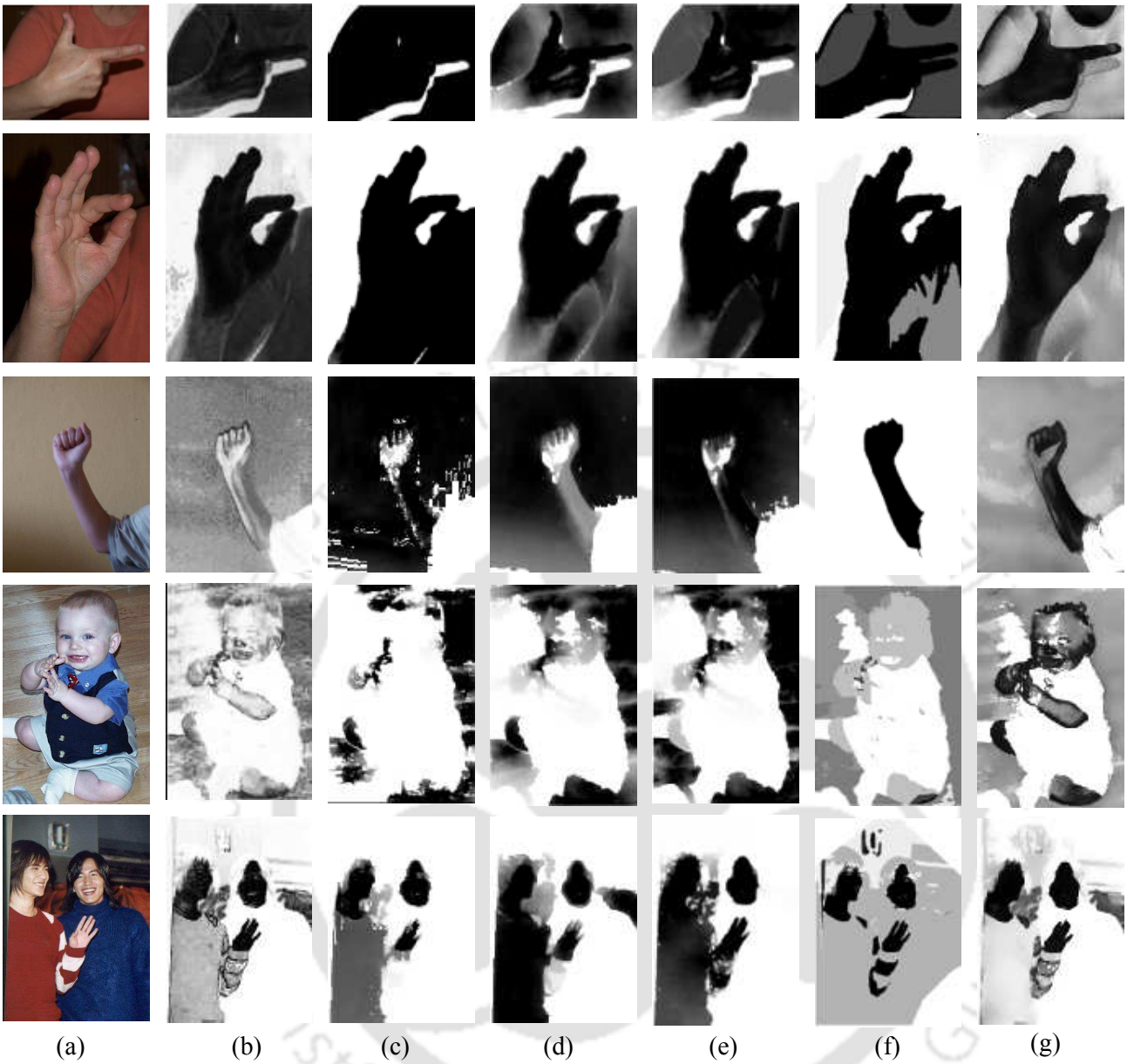


Figure 2.12: Comparative demonstration of skin maps of some sample images: a) Original image, b) Bayes classifier [13], c) FASS [14], d) DSPF [9], e) SASS [15], f) MMSC [16], and g) Proposed DSM.

colour similarity.

2.4 Summary

In our proposed skin segmentation method, discrimination between skin and non-skin pixels are enhanced with the help of Adaptive Discriminant Analysis (ADA). The proposed ADA adapts to scene characteristics and extracts image specific discriminant feature for skin and

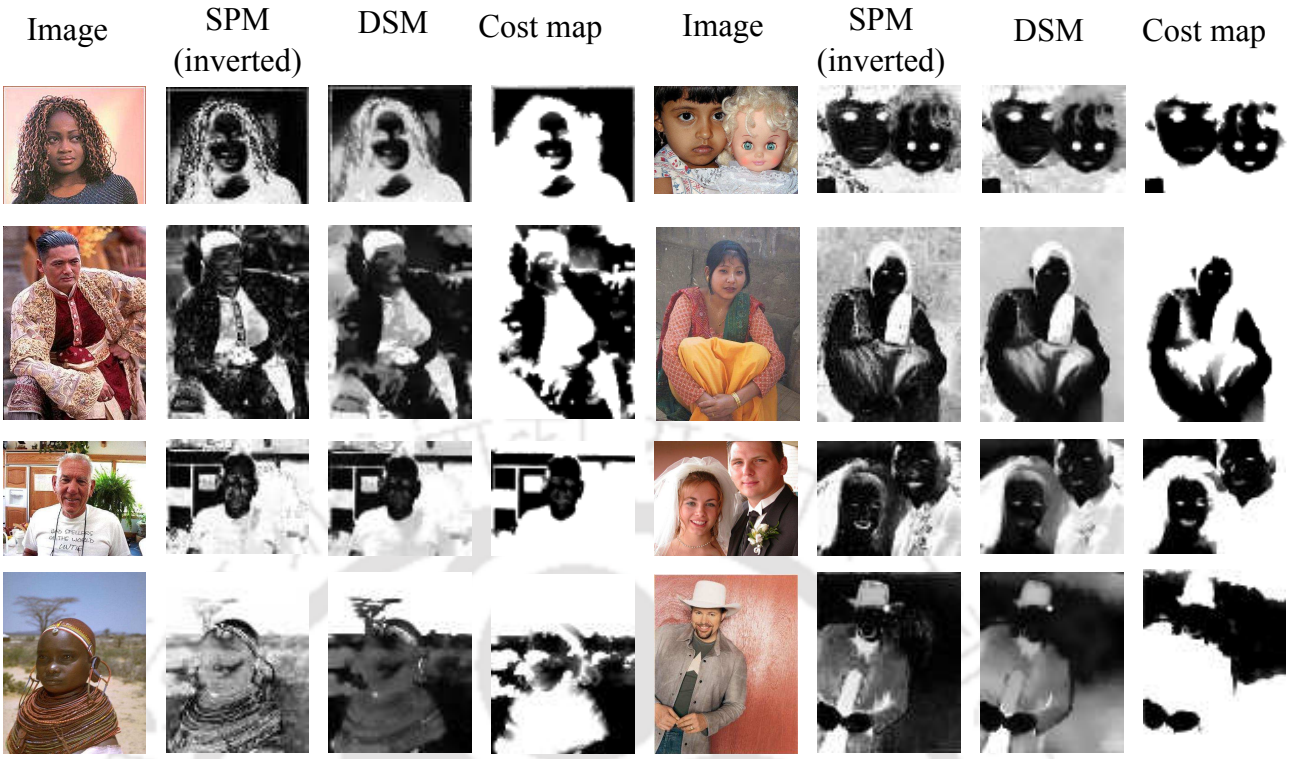


Figure 2.13: Unsatisfactory cost maps for some sample images selected from ECU dataset.

non-skin pixels. Additionally, a dynamic region growing (DRG) method is employed in our method. The DRG allows skin regions to grow dynamically based on a feedback input in the form of false detection. The main advantage of our proposed method is that the most discriminant features are extracted from an image itself, and hence, our method is suitable for different backgrounds. However in some of the instances, skin regions have altogether different colour characteristics due to poor lighting conditions and/or coloured light sources. In these specific situations, SPM fails to identify the skin pixels. Also, the proposed ADA may fail to accurately detect skin regions when the non-skin regions have almost similar colour and texture characteristics as that of skin regions. Figure 2.13 shows some of such cases where the proposed method does not produce satisfactory results.

3

Detection of Skin using Image Pixel Distribution Information

The previous method can not perfectly handle non-uniform illumination as improper illumination makes skin regions to appear darker than its actual tone. Also, the derivation of an accurate discriminative space map of an image is possible only if the largest skin-like coloured region in an image belong to skin. To overcome these limitations of the previous method, a new skin detection method is proposed which is based on a local skin model adaption scheme by utilizing image pixel distribution information. The distribution image pixels in a given colour space is approximated by a Gaussian mixture model (GMM), which is termed as image distribution model (IDM). In this method, a local skin distribution model (LSDM) and a local background distribution model (LBDM)are derived by finding the similarities of the Gaussian components of IDM to a given reference skin distribution model. The reference skin model is derived from a set of facial skin pixels, and it is termed as facial skin distribution model (FSDM). Subsequently, a local skin probability map (LSPM) is derived by using the LSDM and the LBDM. Finally, the LSPM is fused with a skin probability map obtained from a global skin model by following a fusion rule, and a fusion-based skin probability map (FSPM) is derived. The proposed DRG algorithm in ***Chapter 2*** *is applied on the FSPM for additional reduction in detection errors. Experimental results show that the proposed FSPM can better discriminate skin regions from non-skin regions as compared to the state-of-the-art methods.*

3.1 Introduction

The accuracy of colour-based skin segmentation algorithms are limited due to the presence of some colours in the background, which are similar to human skin colour, and poor illumination conditions. Skin like-colours in the background increases false positive detection. On the other hand, poor illumination may altogether change the chrominance properties of the actual skin colours, and the actual skin colour will be different from the image colour. Many authors like [99, 120–122] converted the colour space from RGB space to another space, dropped the luminance component, and used only the chrominance components in order to compensate the brightness variations in the image. However, Storring *et al.* [123] showed that the skin reflectance locus and the illuminant locus are directly related. This means that the perceived colour depends on scene illumination.

To overcome these problems, many researchers proposed local adaptation schemes for a global skin detection model by utilizing local chromatic and/or textural informations. It is observed that facial skin colour resembles the overall skin tone of a person. Motivated by this fact, some researchers extracted pixels of the facial regions in and image and used them as a reference skin tone for local adaptation of the global skin detection model. For example, Kawulok *et al.* [19] proposed a dynamic skin model, where the global pixel statistics are fused with local statsitics of facial skin pixels. Yogarajah *et al.* [20] proposed a dynamic thresholding-based pproach by using chromatic properties of facial skin pixels. Tan *et al.* [10] fused a smoothed colour histogram and a Gaussian skin model with the help of facial skin pixels as reference. Kawulok *et al.* [21] used a local skin distribution model derived from facial skin pixels to obtain seed regions for their propagation-based skin segmentation method. However, skin appearance over a body changes on account of non-uniform illuminations. Therefore, a skin model obtained only from the facial skin pixels cannot generalize the overall skin colour distribution of a person due to insufficient training skin samples.

In view of the above-mentioned issues relating to accurate skin region detection, a novel skin detection algorithm is proposed by utilising the information of the distribution of image pixels. We considered the principle that an image can be clustered into a number of Gaussian distributed clusters [124]. So, the pixel distribution can be modelled as a mixture of Gaussian functions. The image pixel distribution model is termed as *image distribution model* (IDM). So, the Gaussian distributed clusters of skin pixels should be statistically closer to a reference skin

pixel distribution in a given colour space. The reference skin pixel cluster could be either obtained from a skin model derived from a set of skin patches belonging to different people or from the facial pixels of a person present in the image. Subsequently, a new Local Skin Probability Map (LSPM) is derived by measuring the similarity between IDM and the reference skin pixel distribution model. Finally, a fusion-based probability map is derived by combining a global skin probability map (GSPM) with the proposed LSPM. Subsequently, a the proposed dynamic region growing (DRG) method is applied in order to improve the segmentation accuracy. The proposed method is discussed in details in the section to follow.

3.2 Proposed Method

In our proposed method, the image pixel distribution model is coupled with some reference skin pixels to capture the variations of local skin colours. The local skin detection model is then fused with a skin detection model derived from some globally obtained samples. Performance-wise, a fusion-based model can give more accuracy as compared to either global or local skin models. The proposed skin segmentation algorithm has mainly four basic steps, and all these steps are described in the following sections.

3.2.1 Global skin probability map

To derive a skin probability map (SPM), two sets of training samples – one for skin and another for non-skin class are required. The training samples are obtained from a set of randomly selected images from a standard dataset. That is why these samples are termed as "global" samples, and the corresponding SPM is termed as "Global Skin Probability Map" (GSPM). Derivation of GSPM involves derivation of two global likelihood functions (GLFs) for skin and non-skin pixels. The GLFs are obtained through training of two GMMs of the samples of Skin (S) and Non-skin (NS) pixels. The GLF of skin pixels can be modelled using GMM as:

$$P(\mathbf{X}|S) = \sum_{i=1}^{K_{gs}} \omega_i^{gs} N\left(\mu_i^{gs}, \Sigma_i^{gs}\right) \quad , \tag{3.1}$$

where, ω_i^{gs} is the prior probability weight, K_{gs} is the number of Gaussians used for GMM, μ_i^{gs} is the mean, and Σ_i^{gs} covariance matrix of the i^{th} Gaussian $G_i\left(\mu_i^{gs}, \mathbf{\Sigma}_i^{gs}\right)$. In this expression, $\mathbf{X}$

is the colour component vector. In this thesis, this model is termed as *global skin distribution model* (GSDM), which is expressed as,

$$G^{gs} \equiv P(\mathbf{X}|S), \tag{3.2}$$

Similarly, the GLF for non-skin pixels is given as:

$$P(\mathbf{X}|NS) = \sum_{k=1}^{K_{gb}} \omega_k^{gb} N\left(\mu_k^{gb}, \Sigma_k^{gb}\right) \quad , \tag{3.3}$$

where, K_{gb} is the number of Gaussians in the non-skin GMM. In this thesis, this model is termed as global background distribution model (GBDM), which is expressed as,

$$G^{gb} \equiv P(\mathbf{X}|NS), \tag{3.4}$$

The Global skin probability map (GSPM) of an image is given by:

$$GSPM(\mathbf{X}) = \frac{P(S)P(\mathbf{X}|S)}{P(S)P(\mathbf{X}|S) + P(NS)P(\mathbf{X}|NS)} \quad . \tag{3.5}$$

where, $P(\mathbf{X}|S)$ and $P(\mathbf{X}|NS)$ are the Skin (S) and Non-skin (NS) likelihoods respectively. $P(S)$ and $P(NS)$ are skin and non-skin priors calculated from the training skin and non-skin pixels.

3.2.2 Image pixel distribution model

Image pixel distribution gives an insight of an image characteristics. We assume that pixels of an image group into multiple overlapped Gaussian distributed clusters. So, the pixel distribution $h(\mathbf{X})$ of an image can be modelled as a mixture of Gaussians with maximum number of components K_{max}. Here, $\mathbf{X} = [\mathbf{R}, \mathbf{G}, \mathbf{B}]'$ is the colour vector. Initialization of the GMM is done with the help of k-means algorithm, and the cluster means ($\mu^{\mathbf{I}}$) are obtained. The covariance matrices ($\mathbf{\Sigma}^I$) for all the Guassians are assumed to be identity matrices. Also, the prior Gaussian weights (ω^I)) are assumed to be equal. Therefore, we can model the image pixel distribution as:

$$h(\mathbf{X}) = G(\mu^I, \mathbf{\Sigma}^I, \omega^I) \quad , \tag{3.6}$$

Algorithm 1: Proposed EM-algorithm

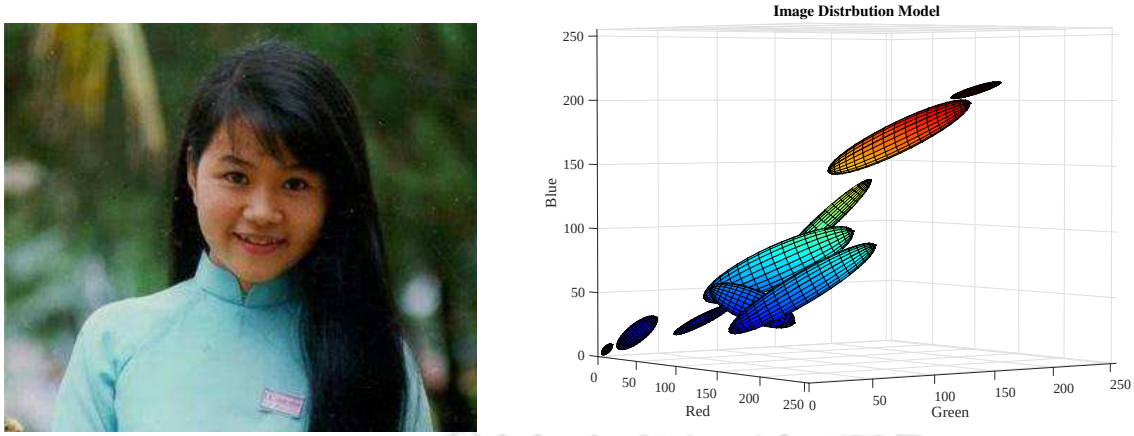
Data: $\mathbf{X_i}$ where, $i = 1, .., M \times N$
Result: $\mu^I, \mathbf{\Sigma}^I, \omega^I, K_I$
Initialize: $\mu^I, \mathbf{\Sigma}^I, \omega^I$
Set: Maximum number of iterations $iter_{max}$
Set: $K = K_{max}$
for $1 \leq i \leq iter_{max}$ **do**
  E-Step:
  Compute the membership weights ω_k^I
  M-Step:
  Compute the parameters $\mu^I, \mathbf{\Sigma}^I, \omega^I$
  for $1 \leq k \leq K$ **do**
    if $\Sigma_k^I = singular$ **then**
      Drop the k^{th} Gaussian
      $K \Leftarrow K - 1$
      $K_I = K$
    else
      Do not update K_I
    end
  end
end
return $\mu^I, \mathbf{\Sigma}^I, \omega^I, K_I$

However in some of the instances, modelling of the pixel distribution of an image using GMM not be suitable due to the occurrence of ill-conditioned co-variance matrices of some clusters. These clusters may have significantly less number of samples. To address this application specific issue, the M-step of the EM-algorithm for GMM is modified to suit our purpose. For this, a singularity check of the co-variance matrices is performed at each of the iteration in the M-step. We performed the singularity check with the help of Statistics and Machine Learning Toolbox of Matlab®. If the co-variance matrix fails the singularity check, then the corresponding Gaussian component is dropped from the mixture. After the convergence of EM algorithm, the final value of Gaussian components K_I is obtained. Hence, the IDM is given as:

$$h(\mathbf{X}) = \sum_{k=1}^{K_I} \omega_k^I N\left(\mu_k^I, \Sigma_k^I\right) \quad . \tag{3.7}$$

3.2.3 Fusion-based skin probability map

This step has two parts – a) Derivation of a local skin and background distribution model, and b) Fusion of a local skin probability map (LSPM) derived from these local distribution



(a) (b)

Figure 3.1: Illustration of GMM modelling of image pixel distribution: a) Original image, b) Image pixel distribution model.

models with the GSPM. The derivation of a local skin distribution model (LSDM) and a local background distribution model (LBDM) requires a reference skin colour distribution model. For this, we follow two approaches – a) selection of GSDM as the reference skin distribution model, and b) derivation of local skin distribution model from a set of pixels belonging to the facial region(s) of the person(s) present in an image. The detailed analysis is given below.

3.2.3.1 FSPM derivation by using GSDM

In this approach, an LSDM and an LBDM are derived from the IDM by measuring its similarity with the GSDM as shown in Figure 3.2. It is observed that the clusters of IDM, which have skin pixels, are statistically closer to the GSDM. Therefore, The inverse of the inter-cluster distance between the IDM and the GSDM can be used as a skin similarity index for the Gaussian components of IDM. The inverse function is more sensitive to the distance changes in lower distance range as compared to higher distance range. On the other hand, the logarithm function is less sensitive to the variation of argument in a higher range as compared to a lower range. So, taking the logarithm of the distance values gives a non-skin similarity index, which shows approximately complementary variation in the skin similarity index variation. Therefore, the skin similarity index ($d_{S,i}^{GI}$) and the non-skin similarity index ($d_{NS,i}^{GI}$) of the i^{th}

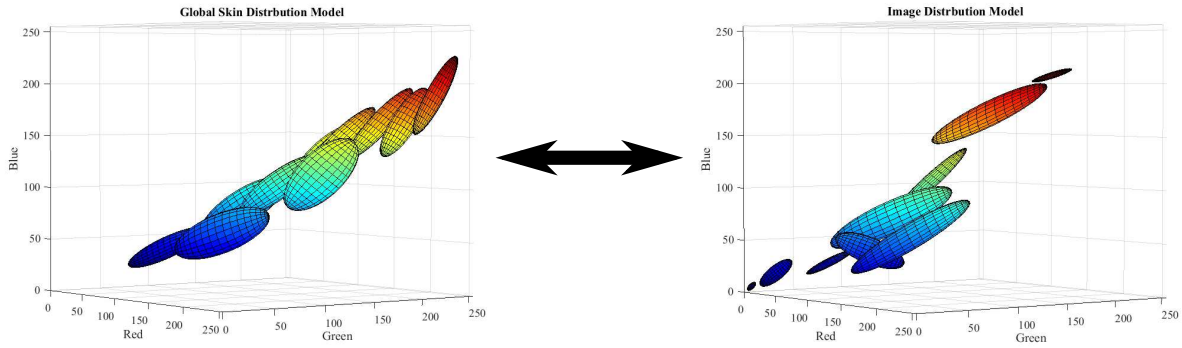


Figure 3.2: Similarity match between GSDM and IDM.

Gaussian component G_i^I are expressed as:

$$\begin{aligned} d_{S,i}^{GI} &= \left[\min_{1\le j\le K_{gs}} \left\{D\left(G_i^I, G_j^{gs}\right)\right\}\right]^{-1} \\ d_{NS,i}^{GI} &= \log\left[1 + \min_{1\le j\le K_{gs}} \left\{D\left(G_i^I, G_j^{gs}\right)\right\}\right] \end{aligned} \quad , \tag{3.8}$$

where,

$$D\left(G_i^I, G_j^{gs}\right) = \left(\mu_i^I - \mu_j^S\right)' \cdot \left(\mathbf{\Sigma}_j^{gs}\right)^{-1} \cdot \left(\mu_i^I - \mu_j^{gs}\right) \quad . \tag{3.9}$$

Here, the superscript GI denotes similarity between GSDM and IDM. An LSDM (h_S) should highlight the skin pixel probability, whereas the LBDM (h_{NS}) should highlight non-skin pixel probability. Hence, the local distribution models for skin and non-skin pixels of an image are expressed as:

$$\begin{aligned} h_S(\mathbf{X}) &= \sum_{i=1}^{K_I} \lambda_i^S \cdot N\left(\mu_i^I, \mathbf{\Sigma}_i^I\right) \\ h_{NS}(\mathbf{X}) &= \sum_{i=1}^{K_I} \lambda_i^{NS} \cdot N\left(\mu_i^I, \mathbf{\Sigma}_i^I\right) \end{aligned} \quad , \tag{3.10}$$

where,

$$\lambda_i^S = \frac{d_{S,i}}{\sum_{i=1}^{K_I} d_{S,i}} \text{ and } \lambda_i^{NS} = \frac{d_{NS,i}}{\sum_{i=1}^{K_I} d_{NS,i}} \quad . \tag{3.11}$$

Therefore, Local SPM (LSPM) of an image is given by:

$$LSPM_{NoFace}(\mathbf{X}) = \frac{P(S)h_S(\mathbf{X})}{P(S)h_S(\mathbf{X}) + P(NS)h_{NS}(\mathbf{X})} \quad . \tag{3.12}$$

Hence, the Fusion-based SPM (FSPM)[1] is obtained as:

$$FSPM_{NoFace}(x) = \rho LSPM_{NoFace}(x) + (1-\rho)GSPM(x) \quad , \tag{3.13}$$

where, $\rho \in [0,1]$. However, the applicability of this approach is limited in the presence of skin-like colours in the background. Skin-like background colours may show significant similarity with actual skin colours. This may result in a large amount of false acceptance error. It is evident that facial skin colour resembles the skin colour of a person. Therefore, the similarity measurement should be performed between the IDM and a model describing facial skin pixel distribution also termed as facial skin pixel distribution model (FSDM). This helps in rejecting the skin-like coloured background regions more effectively.

3.2.3.2 FSPM derivation by using FSDM

The proposed fusion-based skin modelling locates the skin pixels in the facial region. At first, the face is located using the algorithm proposed by Viola and Jones [125]. The distribution of reference skin pixels extracted from the face region is modelled as a single multi-variate Gaussian function $G^f \equiv N\left(\mu^f, \mathbf{\Sigma}^f\right)$. If the face and the other body parts are uniformly illuminated, then the skin pixels can be detected with the help of a local face model derived from the skin pixels of the facial region. But in practice, the local face model may fail to detect the apparent colour variations in the skin regions of different body parts due to non-uniform illumination. Our observation is that in spite of small local colour variations, the overall chromatic properties of the skin regions do not change significantly. So, the clusters of IDM with actual skin pixels will be statistically closer to the local face model. Subsequently, a local skin probability map ($LSPM_{Face}$) is derived using the IDM and its similarity with FSDM as shown in Figure 3.3. We use the following notation in the rest of this chapter for simplicity, $G_i^I \equiv N\left(\mu_i^I, \mathbf{\Sigma}_i^I\right)$. Hence, the skin similarity index ($d_{S,i}^{FI}$) and the skin dissimilarity index ($d_{NS,i}^{FI}$) of the i^{th} Gaussian component

[1]This work has been published in *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP)*, ACM 2016 (Refer item 1 in Page 117 for details)

G_i^I is expressed as:

$$d_{S,i}^{FI} = \frac{1}{D\left(G_i^I, G^f\right)} \text{ and } d_{NS,i}^{FI} = \log\left[1 + D\left(G_i^I, G^f\right)\right] \quad , \tag{3.14}$$

where,

$$D\left(G_i^I, G^f\right) = \left(\mu_i^I - \mu_j^f\right)' \cdot \left(\mathbf{\Sigma}_j^f\right)^{-1} \cdot \left(\mu_i^I - \mu_j^f\right) \quad . \tag{3.15}$$

Here, the superscript FI denotes Face-Image similarity. The proposed skin and skin dissimilarity indexes are used to derive the local distribution models for both the skin and the non-skin pixels of an image. The LSDM for skin pixels (h_S^F) and the LBDM for non-skin pixels (h_{NS}^F) are expressed as:

$$h_S^f(X) = \sum_{i=1}^{K_I} \eta_i^S \cdot N\left(\mu_i^I, \Sigma_i^I\right) \text{ and } h_{NS}^f(X) = \sum_{i=1}^{K_I} \eta_i^{NS} \cdot N\left(\mu_i^I, \Sigma_i^I\right), \tag{3.16}$$

where,

$$\eta_i^S = \frac{d_{S,i}^{FI}}{\sum_{i=1}^{K_I} d_{S,i}^{FI}} \text{ and } \eta_i^{NS} = \frac{d_{NS,i}^{FI}}{\sum_{i=1}^{K_I} d_{NS,i}^{FI}} \quad . \tag{3.17}$$

Therefore, the Local SPM (LSPM) of an image is given by:

$$LSPM_{Face}(x) = \frac{P(S)h_S^f(\mathbf{X})}{P(S)h_S^f(\mathbf{X}) + P(NS)h_{NS}^f(\mathbf{X})} \quad . \tag{3.18}$$

The LSPM generally produces more discrimination between the skin and the non-skin regions. But, it also assigns higher probability values to the skin regions or colours similar to facial region. So, there is a chance of more false detection. That is why, the fusion should be done in such a way that the overall detection error is low. Also, the face may be inadequately illuminated. For such specific situations, the GSPM will have less influence on the overall skin model. Keeping all these considerations in view, we proposed a fusion rule to derive the

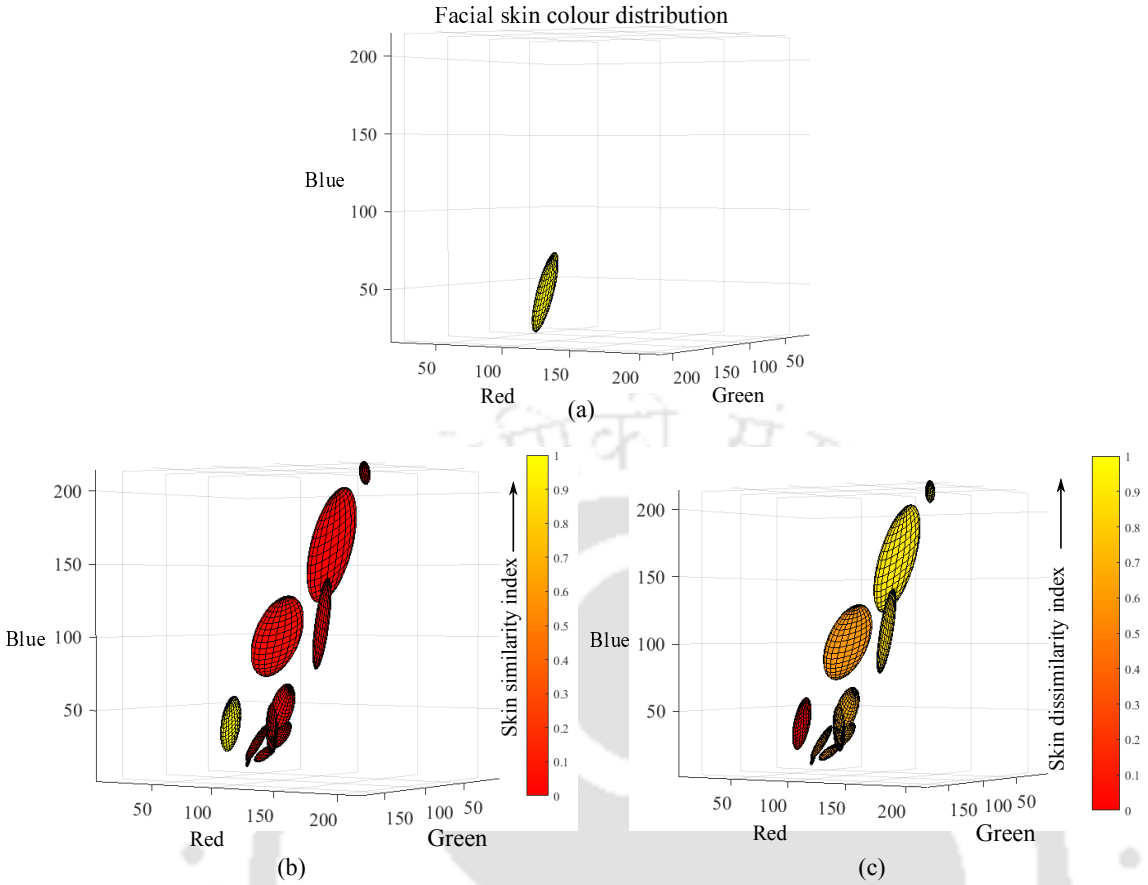


Figure 3.3: Similarity in Face-Image model: a) Skin model derived from facial skin pixels, b) IDM with each cluster coloured in accordance with the skin similarity index, and c) IDM with each cluster coloured in accordance with skin dissimilarity index. In the colourmap, Red represents smallest index value, whereas Yellow represents highest index value. The Yellow-coloured cluster in (b) is closest to the skin colour.

proposed Fusion-based SPM ($FSPM_{Face}$) as follows:

$$\begin{aligned} &\text{If } GSPM^{f}_{avg} \geq \sigma, \\ &\quad FSPM_{Face}(\mathbf{X}) = \min\{GSPM(\mathbf{X}), LSPM_{Face}(\mathbf{X})\} \\ &\text{otherwise,} \\ &\quad FSPM_{Face}(\mathbf{X}) = \alpha \cdot LSPM_{Face}(\mathbf{X}) + (1-\alpha) \cdot GSPM(\mathbf{X}) \end{aligned} \tag{3.19}$$

where, $GSPM^{f}_{avg}$ is the average GSPM value of the face masked pixels, and $\alpha \in [0, 1]$. σ is a thresholding value for GSPM. Apparently, both σ and α should be high enough to minimize the overall detection error. Figure 3.4 shows discrimination between the skin and the non-

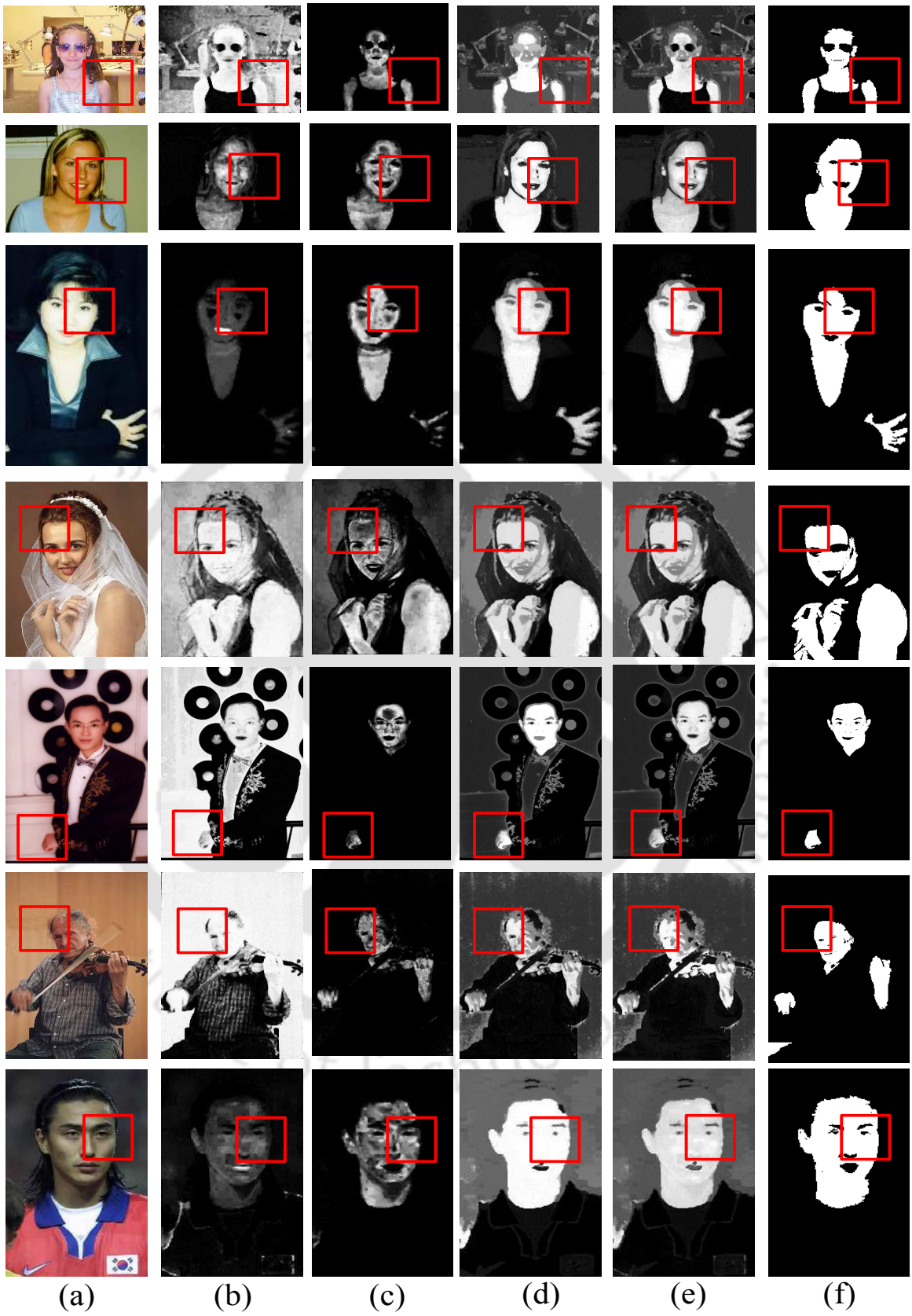


Figure 3.4: FSPM from face: a) Original image, b) GSPM [13], c) SPM derived directly from the pixels of a facial region, d) LSPM derived from face-image model similarity, e) Proposed FSPM, and f) Groundtruth results. The red coloured bounding boxes show the specific regions, where our proposed FSPM gives comparatively better discrimination between the skin and the non-skin regions.

skin regions. It is evident that the proposed FSPM[2] is able to produce more discrimination between the skin and the non-skin regions for different scene characteristics ranging from low illumination to skin-like backgrounds.

The final skin mask for a given image can be easily obtained by direct thresholding of either of the above FSPMs with a suitable threshold. However, direct thresholding may consider some non-skin regions with comparatively higher FSPM values as skin. This increases the chance of false positive error. Therefore, we use the dynamic region growing (DRG) method as proposed in **Chapter 2** to obtain a cost map $C(x)$ for skin regions. The final skin mask is obtained by thresholding the cost map $C(x)$ with a threshold θ_{th}.

3.3 Experimental Analysis

For the validation of our proposed method, four datasets are used: Cambridge hand gesture dataset [68,69], NUS dataset [70], HGR dataset [9, 14, 67], and ECU dataset [65]. The images in Cambridge dataset are taken under varying illumination conditions with mostly non-skin looking backgrounds. A set of 200 images belonging to this dataset are manually annotated for the experimental validation process. The NUS dataset contains hand pose images taken under random background and natural illumination conditions. We selected a set of 500 images from NUS dataset, which are manually annotated. The HGR dataset contains 899 images captured under varying illumination conditions in an indoor environment with some images having skin-like coloured backgrounds. The ECU dataset contains two sets of images (each containing 2000 images), which are captured in uncontrolled natural environments. To obtain the GSPM, skin and background patches from one set, while the other is used for experimental validation. The optimum number of Gaussian components for the GSDM *i.e.* (K_S) and the GBDM *i.e.* (K_{NS}) are obtained using Bayesian information criterion (BIC). In our case, the optimum value of the K_S is found to be 6, and that of the K_{NS} is 9.

In this experimental validation process, four measures are selected for evaluation, namely - *detection accuracy (Acc.)*, *false positive error rate* (δ_{fp}), *false negative error rate* (δ_{fn}), *total detection error rate*(δ_t), where $\delta_t = \delta_{fp} + \delta_{fp}$. All the detection error values are obtained by thresholding the cost map with a threshold θ_{th}. Here, θ_{th} corresponds to the minimum δ_t.

We also analyse the dependency of δ_t and computational time on the maximum number of

[2]This work has been published in *Pattern Recognition Letters, 2017.* (Refer item 1 in Page 117 for details)

components K_{max} of the Gaussian mixture model. For this, δ_t value is calculated for different values of K_{max}. It is observed that, δ_t decreases very slowly for $K_{max} > 11$. On the other hand, an increase of K_{max} increases the computational time. It is observed that the average computation time for an image almost linearly increases with K_{max}. Keeping these two conflicting requirements in view, the optimum value of K_{max} is fixed at 11. It is found thatthe DRG produces an optimum result for FSPMs at $P_{s,max} = 0.8$.

3.3.1 Experimental validation for $FSPM_{NoFace}$

It is observed that GSPM gives a comparatively better accuracy in presence of background similarity but, performs poorly in case of poor illumination. On the other hand, the local skin model performs well in presence of poorly illuminated conditions but performs poorly in presence of background colour similarity. Hence, the parameter ρ is selected to be 0.5 for the validation of $FSPM_{NoFace}$, and kept fixed for all the datasets.

At first, we analysed the effect of β on the overall detection error. For this, detection errors are calculated with varying β values for the different dataset. A plot illustrating δ_t vs β is obtained for different datasets as shown in Figure 3.5. It is observed that δ_t becomes the lowest for all the dataset except for the Cambridge dataset at $\beta \approx -1$. For HGR and NUS dataset, δ_t slowly increases when $\beta > -1$, whereas it remains almost constant for ECU dataset. Cambridge is the only dataset whose total detection error keeps reducing with decreasing β owing to very little skin-like colour present in the background. This makes region growing less restricted. However, images of Cambridge dataset belong to a specific situation, where illumination is non-uniform and there are no skin-like colours present in the background. In general, skin-like colours could be present in the background. Therefore, we choose $\beta = -2$ for the rest of this validation process.

For experimental validation, the proposed method is compared with a set of state-of-the-art methods, such as - Bayes classifier [13], Stacked Autoencoders-based skin classification (SASC) [22], Fast propagation-based skin segmentation (FPSS) [14], Self-adaptive seed selection (SASS) [15], Multi manifold-based skin classification (MMSC) [16], Flexible neural tree (FNT) [119], Random forest-based skin classification (RFSC) [17], CMYK colour space-based skin classification [18]. The comparative results for Cambridge dataset and NUS dataset are given in Table 3.1. For HGR and ECU dataset, comparative results are provided in Table 3.2. It is

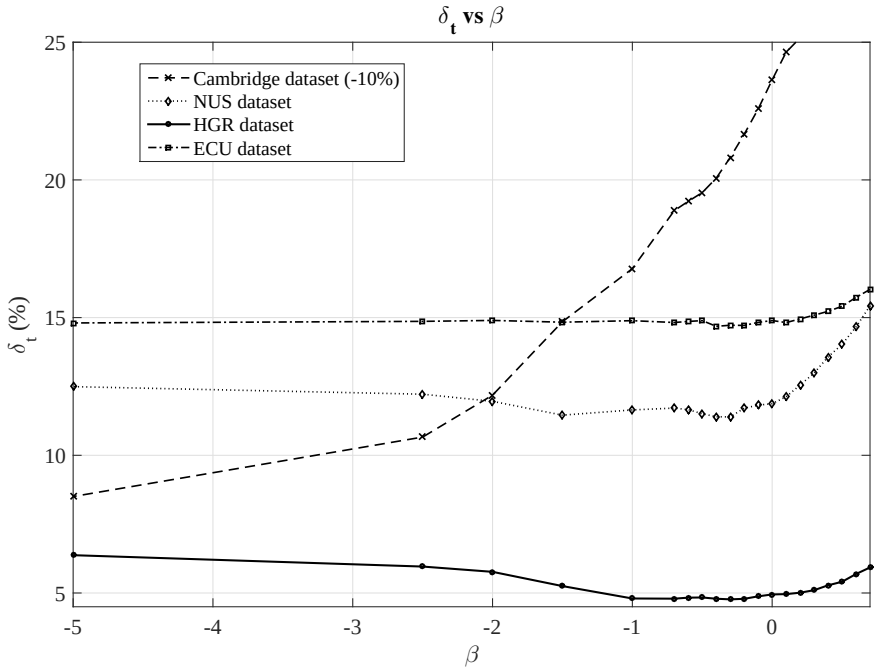


Figure 3.5: δ_t vs β for different datasets

clearly visible that the fusion of image pixel distribution information with a global skin colour model significantly improves the skin detection process for all the datasets. The images of Cambridge dataset are taken under comparatively higher non-uniform illumination conditions than the other datasets. Thus, false negative rate for Cambridge dataset is the highest among all the datasets. Now, region growing-based methods reduce false positive rate at a cost of an increased false negative error rate. So, the direct thresholding-based method produces comparatively less overall detection error rate in this case. The images of NUS and ECU dataset are taken under unconstrained environments with a colour similarity of background to human skin colour in some images. For them, the chance of false positive error rate becomes high in presence of skin-like background colours. So, for these two datasets region growing-based methods detect skin regions with comparatively less overall detection error. The images in HGR dataset are taken under poor but moderately uniform illumination condition. In case of region growing methods, the more is the uniformity in illumination, the better is the growing process. Thus, region growing methods also perform better in case of HGR dataset. The comparative ROC plots for all the datasets are provided in Figure 3.6 and Figure 3.7. The comparative performance analysis is also provided in Figure 3.8. Here, multiple images from each of the datasets are selected to show the detection results. It is clearly observable that the

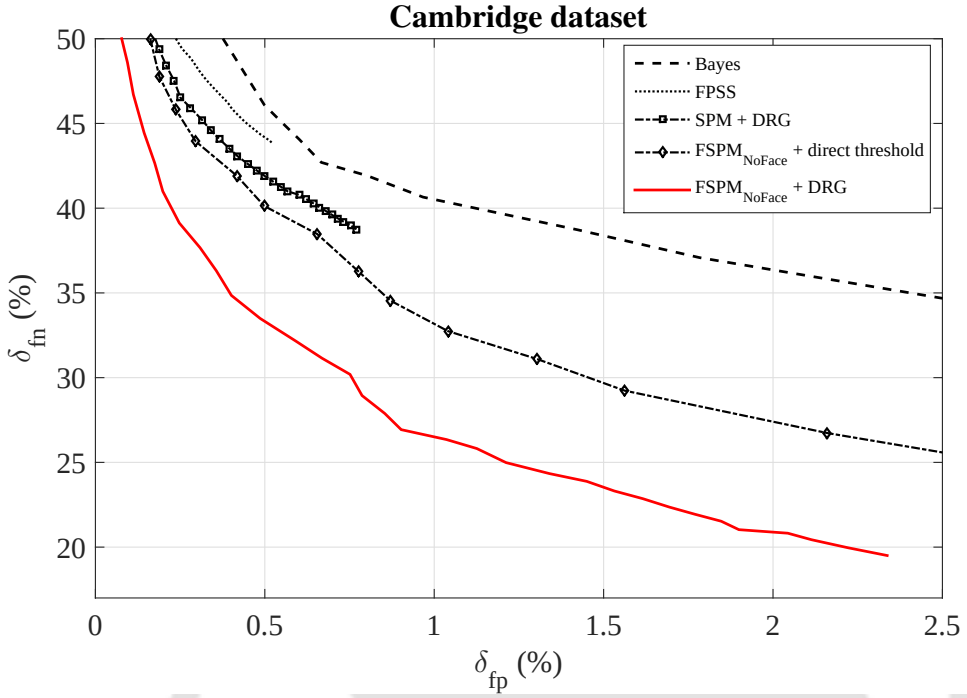
Table 3.1: Comparative analysis of different segmentation methods using Cambridge and NUS dataset

Method	Cambridge database				NUS database			
	$\delta_{fp}(\%)$	$\delta_{fn}(\%)$	$\delta_t(\%)$	**Acc.**	$\delta_{fp}(\%)$	$\delta_{fn}(\%)$	$\delta_t(\%)$	**Acc.**
Bayes classifier [13]	16.63	18.76	35.39	0.8243	11.42	12.65	24.07	0.8837
FPSS[1] [14]	0.52	43.82	44.34	0.8463	9.06	10.95	20.01	0.9047
CMYK colour space [18]	0.41	42.75	43.16	0.8499	16.82	7.49	24.31	0.8474
RFSC [17]	0.61	68.80	69.41	0.7602	3.05	46.83	49.88	0.8950
SPM + **DRG**	0.77	38.76	39.53	0.8615	5.45	11.21	16.67	0.9356
FSPM$_{\text{NoFace}}$ + direct threshold	6.54	10.32	16.86	0.9202	6.83	7.88	14.71	0.9299
FSPM$_{\text{NoFace}}$ + DRG	2.22	19.96	22.18	0.9148	5.85	6.11	11.96	0.9414

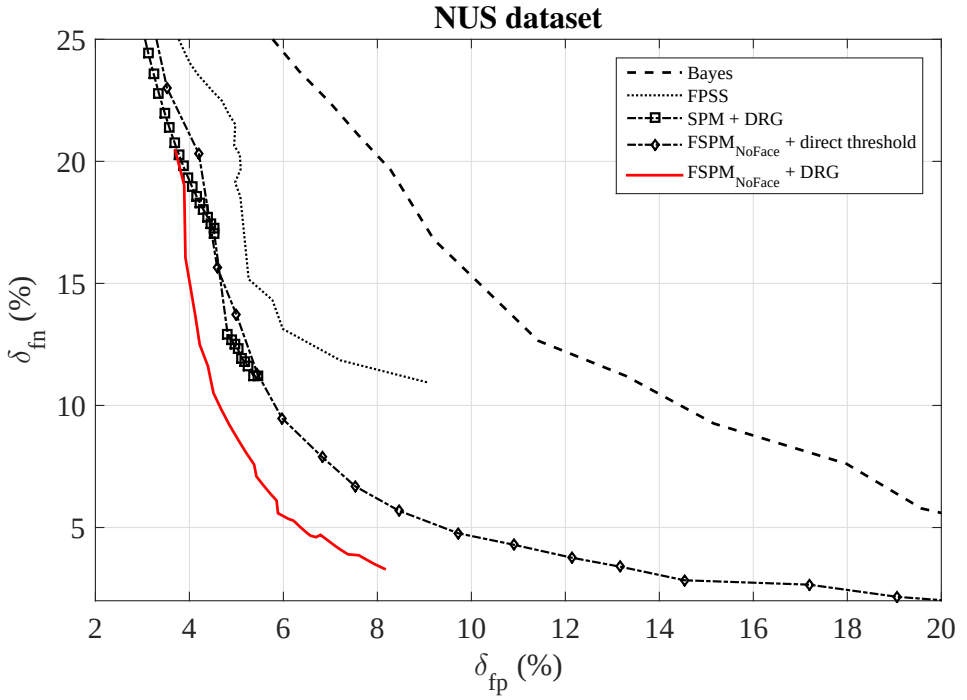
Table 3.2: Comparative analysis of different segmentation methods using HGR and ECU dataset

Method	HGR database				ECU database			
	$\delta_{fp}(\%)$	$\delta_{fn}(\%)$	$\delta_t(\%)$	**Acc.**	$\delta_{fp}(\%)$	$\delta_{fn}(\%)$	$\delta_t(\%)$	**Acc.**
Bayes classifier [13]	5.47	5.15	10.62	0.9470	15.14	9.48	24.62	0.8803
SASC [22]	4.00	20.00	24.00	0.9300	10.00	21.40	31.40	0.8800
FPSS [14]	3.46	5.72	9.18	0.9536	15.05	7.31	22.35	0.8924
CMYK colour space [18]	20.42	2.65	23.07	0.8348	21.79	9.87	31.66	0.8431
RFSC [17]	0.28	36.44	36.72	0.9225	3.05	46.83	49.88	0.8950
FNT[†] [119]	–	–	–	–	9.96	12.86	22.82	0.9053
SASS [15]	3.16	4.57	7.72	0.9611	9.53	7.16	16.68	0.9175
MMSC [16]	1.66	4.41	6.07	0.9692	**4.45**	11.43	15.74	0.9177
SPM + **DRG**	3.30	4.07	7.37	0.9667	8.05	8.02	16.07	0.9198
FSPM$_{\text{NoFace}}$ + direct threshold	2.96	3.29	6.25	0.9710	9.07	9.18	18.25	0.9090
FSPM$_{\text{NoFace}}$ + DRG	2.82	2.94	5.76	0.9724	7.54	7.35	14.90	0.9242

GSPM mostly fails to detect the skin regions of an image under varying illumination conditions. The fusion-based model adapts to the local pixel distribution, and makes the skin pixels more highlighted. It is also observable that the proposed skin detection method is not biassed towards any particular skin tone and performs well for all type of skin colours. However, the proposed FSPM$_{\text{NoFace}}$ derivation relies on relative similarity of IDM pixel clusters with the GSDM pixel clusters. In some situations, some background region may show more skin similarity with GSDM as compared to actual skin regions due to poor illumination. Therefore, the reference skin distribution model should be adaptive to the skin colour of a person present in an image.



(a)



(b)

Figure 3.6: ROC curves for different methods: a) Cambridge dataset, b) NUS dataset.

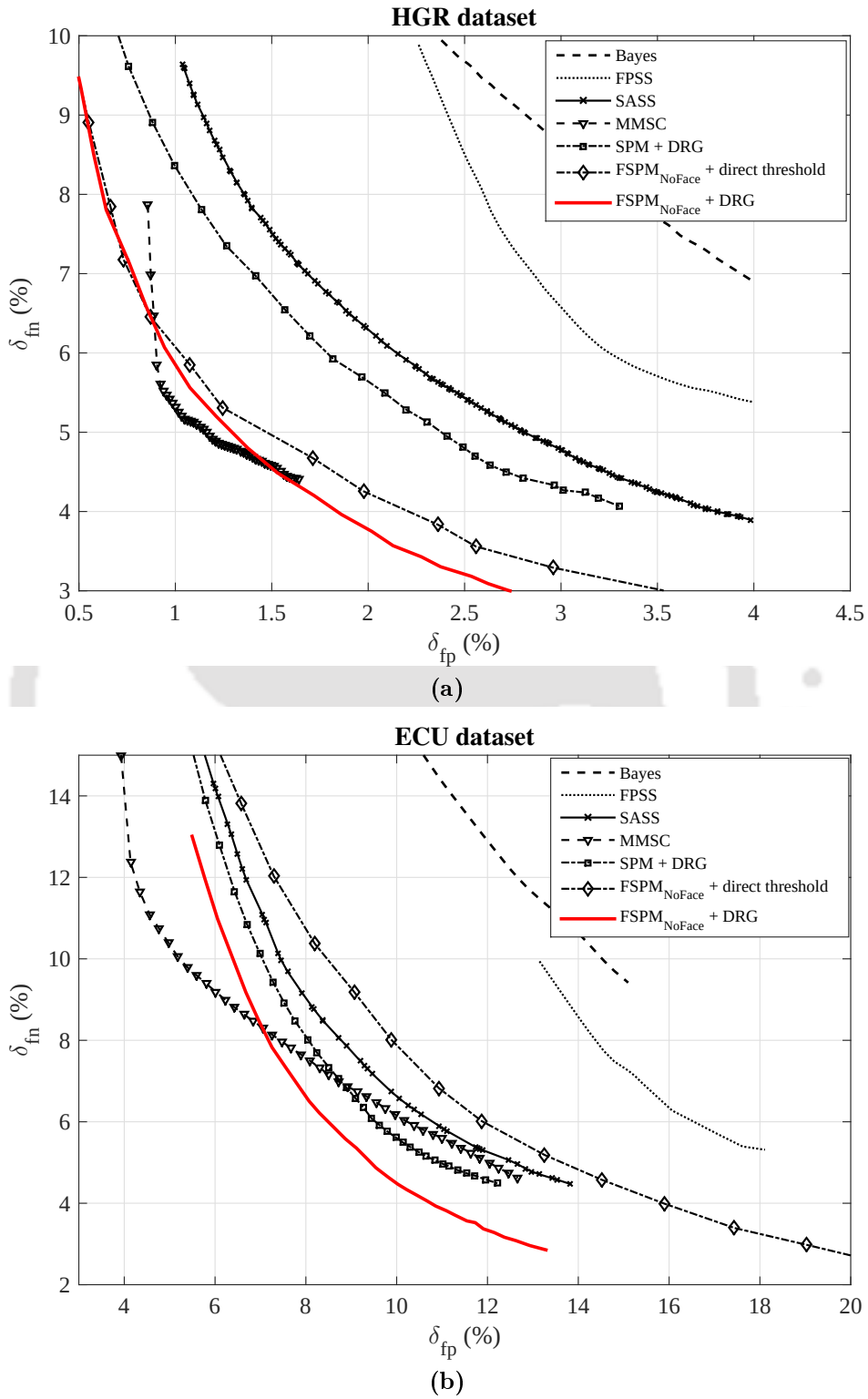


Figure 3.7: ROC curves for different methods: a) HGR dataset, b) ECU dataset.

Figure 3.8: Skin segmentation results : a) original image, b) SPM, c) SPM + direct threshold , d) FPSS [14], e) SPM + DRG , f) Random Forest [17], g) CMYK colour space [18], h) $\text{FSPM}_{\text{NoFace}}$, i) FSPM + direct threshold, j) $\text{FSPM}_{\text{NoFace}}$ + DRG, and k) Ground truth. In the first row, first three images are from Cambridge dataset, next three images are from NUS dataset, next two images are from HGR2b dataset, and final four images are from ECU dataset. Here, white colour represents *true positive*, black colour represents *true negative*, red colour represents *false positive*, and green colour represents *false negative*.

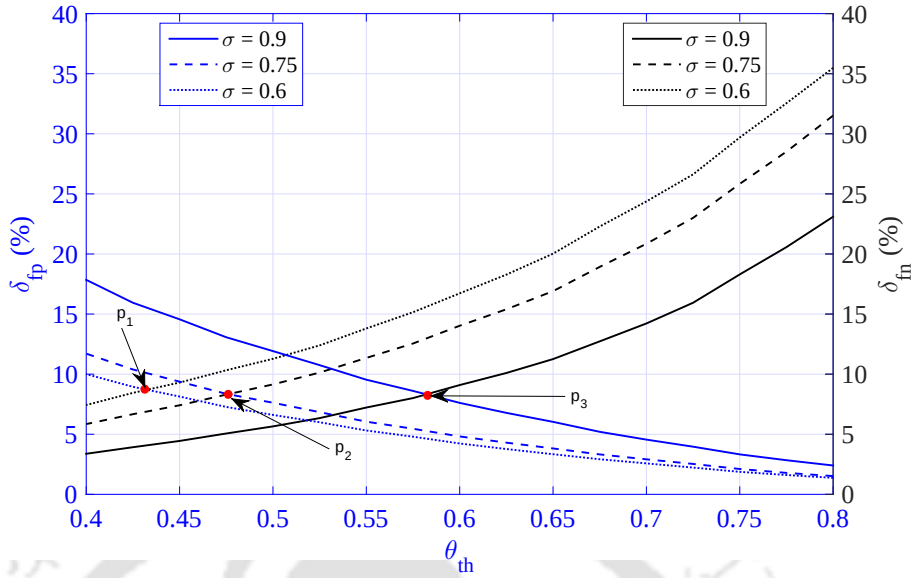


Figure 3.9: Variation of δ_{fp} and δ_{fn} with θ_{th} for different values of σ

3.3.2 Experimental validation for $FSPM_{Face}$

In addition to benchmark methods as mentioned in section 3.3.1, the proposed method is also compared with an additional set of state-of-the-art methods, such as - Dynamic skin model-based skin classification (DSMSC) [19], Adaptive seed-based skin classification (ASSC) [21], Smooth histogram-based skin classification (SHSC) [10], Discriminative Skin-Presence Feature (DSPF) [9], Dynamic threshold–based skin classification (DTSC) [20]. The parameter β is kept fixed at -2 as mentioned in the last section.

The parameter σ maintains a balance between possible false acceptance and false rejection errors. Therefore, it is necessary to obtain an optimum value of σ so that the system can have a more dynamic range of operation without generating significant increase in detection errors. To obtain the optimum value of σ, detection errors are calculated at three instances of σ, such as $\sigma = 0.9$, $\sigma = 0.75$, and $\sigma = 0.6$. Figure 3.9 shows the variation of both δ_{fp} and δ_{fn} w.r.t. θ_{th} for different values of σ. It is observed that $\text{FSPM}_{\text{Face}}$ produces the lowest δ_{fn} for $\sigma = 0.6$ but, gives the highest δ_{fp} for a fixed threshold value θ_{th}. On the other-hand, δ_{fp} is the lowest for $\sigma = 0.9$ but, at the cost of highest δ_{fn}. The points p_1, p_2, and p_3 are the three operating points, where $\text{FSPM}_{\text{Face}}$ produces optimum results with direct thresholding. The proposed FSPM is expected to produce lower false acceptance error during region growing with higher true positive rate. It is found that the proposed $\text{FSPM}_{\text{Face}}$ can produce optimum results at $\sigma \approx 0.75$. Therefore, the

Table 3.3: Comparative analysis of different segmentation methods using ECU dataset

Method	$\delta_{fp}(\%)$	$\delta_{fn}(\%)$	$\delta_t(\%)$	**Acc.**
Bayes classifier [13]	15.14	9.48	24.62	0.8803
SASC [22]	10.00	21.40	31.40	0.8800
DTSC [20]	21.99	9.08	31.07	0.8101
SHSC [10]	10.74	13.36	24.10	0.8882
FPSS [14]	15.05	7.31	22.35	0.8924
CMYK colour space [18]	21.79	9.87	31.66	0.8431
RFSC [17]	3.05	46.83	49.88	0.8950
FNT [119]	9.96	12.86	22.82	0.9053
DSPF [9]	9.31	9.10	18.41	0.9080
DSMSC [19]	8.37	8.38	16.75	0.9171
SASS [15]	9.53	7.16	16.68	0.9175
MMSC [16]	**4.45**	11.43	15.74	0.9177
ASSC [21]	8.41	5.88	14.30	0.9184
SPM + **DRG**	8.05	8.02	16.07	0.9198
$\mathbf{FSPM_{NoFace}}$ + direct threshold	9.07	9.18	18.25	0.9090
$\mathbf{FSPM_{Face}}$ + direct threshold	7.94	8.60	16.54	0.9176
$\mathbf{FSPM_{Face}}$ + **DRG**	5.94	6.15	12.09	0.9384

parameter σ is kept fixed at 0.75 for the validation process.

A comparative analysis of existing skin detection methods with our proposed FSPM-based method is performed. The detailed comparative analysis is shown in Table 3.3. The benchmark methods can be broadly grouped into two classes – methods without face detection and methods using face detection. Methods without face detection include Bayes classifier, SASC, DTSC, FPSS, CMYK colour space-based method, RFSC, FNT, DSPF, SASS, MMSC. On the other hand, skin detection methods, such as SHSC, DSMSC, ASSC use face detection-based reference pixel extraction. The SPM proposed by [13] is a global map for skin colour. On the other hand, the proposed FSPM uses both global statistics of skin colour and also image specific data. So, the proposed skin detection method is more robust to different scene characteristics including illumination variations. Skin detection methods, method proposed by Sawicki *et al.* [18] do not use face for detecting the skin pixels. So, the detection accuracies of these methods are totally dependant on global properties of skin pixels. Among these three methods, Kawulok's method used texture features along with colour characteristics to enhance the discrimination between skin and non-skin regions. Rest of the baseline methods used face detection-based

Figure 3.10: Skin segmentation results : a) Original image, b) SPM, c) Proposed FSPM, d) SPM + direct threshold [13], e) DSMSC [19] + direct threshold, f) DSMSC [19] + FPSS [14], g) DTSC [20], h) SHSC [10], i) ASSC [21], j) DSPF [9], k) $\text{FSPM}_{\text{Face}}$ + direct threshold, l) $\text{FSPM}_{\text{Face}}$ + DRG, and m) Ground truth. Here, white colour represents *true positive*, black colour represents *true negative*, red colour represents *false positive*, and green colour represents *false negative*.

model adaptation. It is evident from the result given in Table 3.3 that skin detection methods, which use face detection-based model adaptation technique perform much better as compared to the methods, which do not use facial regions for model adaptation. The Dynamic thresholding method [20], Fusion of smooth histogram and Gaussian model [10] detect pixels of the facial regions, and the accuracy of skin pixels detection solely depends on the local model only derived from the pixels of facial regions. The method proposed by [9] derives a discriminative space for both colour and texture features without utilising local adaptation. The method performs comparatively better than other listed methods without adaptation. It is evident from Table 3.3 that the proposed fusion-based skin model can produce a better result as compared to the existing methods. The major advantage of selecting $\text{FSPM}_{\text{Face}}$ over $\text{FSPM}_{\text{NoFace}}$ is the former is more user or person specific as compared to the later. Thus, $\text{FSPM}_{\text{Face}}$ can discriminate actual skin regions more accurately in presence of skin like background colours as compared to $\text{FSPM}_{\text{NoFace}}$ (see Table 3.3). Figure 3.10 shows the comparative analysis qualitatively. It is evident from these results that the proposed fusion-based model can more accurately detect the skin pixels in any varying scene characteristics ranging from poor illumination to skin-like backgrounds.

3.4 Summary

In this chapter, a novel fusion-based skin detection method is proposed. The pixel distribution information of an image is used to derive an image specific skin probability map or local SPM. The final skin probability map is obtained by fusing the globally obtained SPM and the local SPM. The fusion of the global and local SPM gives the final fusion-based SPM (FSPM). When skin-like coloured objects are present in the background, the proposed FSPM uses the best parts of both global and local skin models. That is why, our method produces less false positives. On the other hand, global model fails to properly detect the skin regions when the scene illumination is poor. In this situation, the proposed local model uses image pixel distribution information and models the skin pixels distribution more accurately. So, false negative would be less. Experimental results obtained for a standard dataset show that our proposed FSPM can give significantly better results as compared to the state-of-the-art methods, which use only the SPM.

4

Detection of Skin Regions in Videos under Local Colour Deformations

Detection of skin regions in videos is a challenging research problem in presence of uncontrolled illumination conditions. The chromatic appearance of skin regions may change locally due to local shading effects, which may occur due to the motion of different body parts. To address this specific issue, a dynamic adaptation scheme is proposed to detect skin regions which are affected by local colour deformations. The proposed method has two modules – a static module for detection of static skin regions; and a dynamic module for detection of moving skin regions. The static module consists of a facial skin distribution model (FSDM) and a video specific background model. The video specific background model termed as fusion-based background distribution model (FBDM) is obtained by using a local background distribution model (LBDM) and a global background distribution model (GBDM). The LBDM is obtained by considering the similarities of a frame pixel distribution model with the FSDM and the GBDM. Subsequently, the FBDM is derived by using the LBDM and the GBDM. In the dynamic model of the proposed method, an initial moving skin distribution model (MSDM) is obtained by using a GMM from a set of moving skin samples which are obtained by using a modified double frame-difference method. The final MSDM is obtained by performing a filtering procedure based on similarities of the initial MSDM with the FSDM and the FBDM. Finally, the static and the dynamic modules are fused by following a maximisation rule.

4.1 Introduction

Skin detection is an important step for many Human Computer Interaction (HCI) applications. However, various factors make skin detection challenging, such as variations in scene illumination, ethnicity, shadow, background and camera characteristics. One of the major challenges is the effect of illumination variations on skin colour appearance. Skin regions look differently from its actual tone under varying illumination conditions. Illumination variation can occur in two ways – globally and locally. A global variation of illumination occurs when the characteristics of the illuminating source vary with time. On the other hand, local illumination variation occurs when illumination becomes non-uniform over the exposed skin regions. One of the major reason behind the non-uniform illumination is the curved nature of the skin surface. This causes a formation of form shadows on the skin surface. In addition to form shadows, cast shadows may be produced by one of the moving body parts onto another body part. In case of directional light sources, the extent of illumination on a skin patch also depends on its orientation with respect to the light source. In most of the HCI applications, the occurrence of local illumination variations is more frequent and common as compared to global illumination variations.

In recent years, many research works [9,16,22,96] have been reported on skin detection under unconstrained environments for images. However, only few research works have dealt with skin detection in videos under varying illumination conditions. Soriano *et al.* [97] investigated the effects of static but non-uniform illumination on skin colour appearance. Sigal *et al.* [98] proposed a dynamic histogram adaptation model based on a second order Markov model. Here, they assumed that the illumination changes gradually and globally. Habili *et al.* [99] used motion information alongwith colour information to detect skin regions in a video without considering the presence skin-like colous in background. Han *et al.* [101] proposed a skin segmentation and tracking algorithm based on Support Vector Machine (SVM) active learning. In their methoid, the SVM needs to be re-learned at every frame to handle the varying illumination conditions. Liu *et al.* [102] proposed a dynamic skin detection algorithm for videos by utilizing a face detection-based model update scheme for globally varying illumination conditions.

It is evident from the literature survey that the effects of local shading or local colour deformation of skin regions during the skin detection in videos is not fully explored. To address this specific issue, a dynamic adaptation scheme for the detection of skin regions having local colour

deformations is proposed. The proposed method relies on conventional Bayesian classifier. It has two modules– a static module for the detection of static skin regions; and a dynamic module for the detection of moving skin regions. The static module has two components– a skin distribution model for static skin regions and a background distribution model. The static skin distribution model is derived by using a set of facial skin samples of initial frames of a video, and the skin model is termed as *facial skin distribution model* (FSDM). The background distribution model for a video needs to be self-adaptive to the video background characteristics. To obtain a self-adaptive background model for a video, first, a *global background distribution model* (GBDM) is obtained from a set of randomly collected background samples. However, skin-like colours could be present in the background of a video frame. Therefore, a *local background distribution model* (LBDM) needs to be derived for the video frame. To obtain the LBDM, we follow a similarity match-based algorithm by utilizing pixel distribution information of video frames [23]. The similarity between two distribution functions can be measured by using Bhattacharyya distance [126]. The local colour deformations in skin may create multiple modes in the skin distribution. The Bhattachryya distance metric treats those modes as different distribution functions and produces finite distances accordingly. Therefore, a new distance metric is needed to discriminate the distribution functions belonging to similar regions of an image (*e.g.*, skin regions) and the distribution functions belonging to dissimilar regions of an image (*e.g.*, background regions). In this paper, we proposed a modification to the Bhattachryya distance in such a way that it produces lesser distance between two distribution functions than the original Bhattacharyya distance (OBD) if both the distributions belong to skin regions. The LBDM is derived from the FSDM and the GBDM by using the modified Bhattacharyya distance (MBD) as a distance metric. The final background model is obtained by fusing the LBDM and the GBDM, and it is termed as *fusion-based background distribution model* (FBDM). The main component of the dynamic model is a moving skin distribution model (MSDM). The MSDM illustrates the distribution of pixels belonging to moving skin regions having chromatic deformations. The moving skin regions are detected by using a modified double frame-difference method. An initial distribution model for these moving skin regions is obtained by using a GMM. In addition to the actual skin regions, some of the background regions can also be falsely detected as skin regions during the moving object detection process. The final MSDM is obtained by performing a filtering procedure based on similarities of the initial moving skin

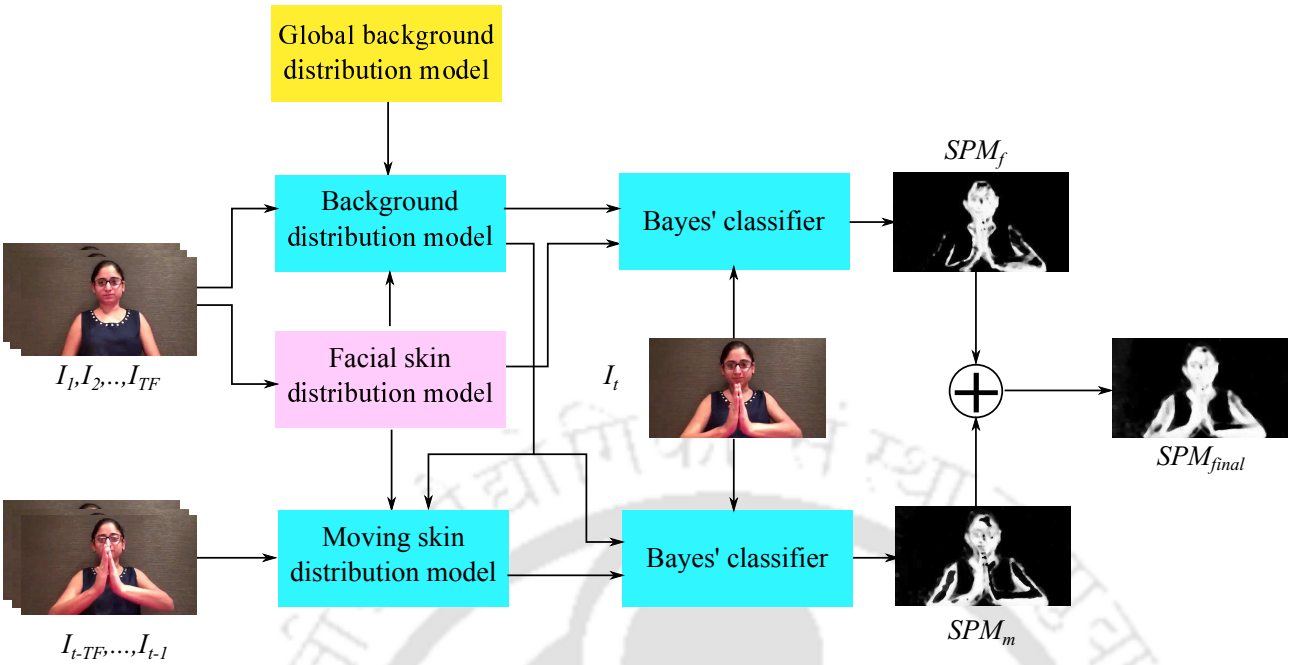


Figure 4.1: Block diagram of the proposed method.

model with the FSDM and the FBDM. Finally, the static and the dynamic modules are fused by following a maximization rule. The proposed method is elaborately discussed in the following sections.

4.2 Proposed Method

The block diagram of the proposed method is shown in Fig. 4.1. The proposed skin detection algorithm has two modules – a static module (FSDM and FBDM) and a dynamic model (MSDM). The background distribution model as shown in Fig. 4.1 is common for both the modules. Each of these modules give one SPM, which are fused together to obtain the final skin probability map for a video frame. The proposed skin detection method is explained in detail in the following sections.

4.2.1 Static module

4.2.1.1 Facial skin distribution model

Facial skin tone can be used as a reference skin tone of a person. At first, Viola and Jones algorithm [125] is applied over a set of initial frames $\mathbf{I}_{TF} = I_1, I_2, .., I_{N_{TF}}$ to locate faces in the

frames. Here, N_{TF} is the number of initial training frames. A set of sample pixels is obtained from the localised facial regions. The distribution of reference skin pixels extracted from the facial regions or the FSDM is modelled as a single multivariate Gaussian function G^f as:

$$G^f = N\left(\mu^f, \mathbf{\Sigma}^f\right) \tag{4.1}$$

4.2.1.2 Background Distribution model

The proposed fusion-based background distribution model has two components – a global background distribution model (GBDM) and a local background distribution model (LBDM). To obtain the GBDM, a set of sample background pixels obtained from a standard dataset is used. The global background model is expressed as:

$$G^{gb} = \sum_{k=1}^{K_{gb}} \omega_k^{gb} N\left(\mu_k^{gb}, \Sigma_k^{gb}\right) \tag{4.2}$$

In order to obtain the LBDM, a local distribution model adaptation scheme [23] is followed. At first, a pixel distribution model G^I for pixels of a set of initial frames $\mathbf{I}_{TF}$ is derived by using a Gaussian Mixture Model (GMM). The expression for G^I is given as:

$$G^I = \sum_{k=1}^{K_I} \omega_k^I N\left(\mu_k^I, \Sigma_k^I\right) \tag{4.3}$$

The number of Gaussian components K_{gb} and K_I are selected by using the method given in [23].

The model G^I gives a colour distribution of frame pixels, which include both skin and non-skin pixels. The Gaussian components of G^I condescending to the real skin regions, should be statistically more similar to FSDM, and they should more dissimilar to the Gaussian components of the GBDM. The Bhattacharyya distance is a well-known metric to find distance between two probability distribution functions. The closed form expression for Bhattacharyya distance between two multivariate Gaussian distribution functions is given by:

$$\begin{aligned} d_{Bh}\left(G_i, G_j\right) =& \frac{1}{8}(\mu_i - \mu_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2}\right]^{-1} (\mu_i - \mu_j) \\ &+ \frac{1}{2} \ln \frac{\left|\left(\Sigma_i + \Sigma_j\right)/2\right|}{\sqrt{\left|\Sigma_i\right| \cdot \left|\Sigma_j\right|}}, \ \forall i, j \end{aligned} \tag{4.4}$$

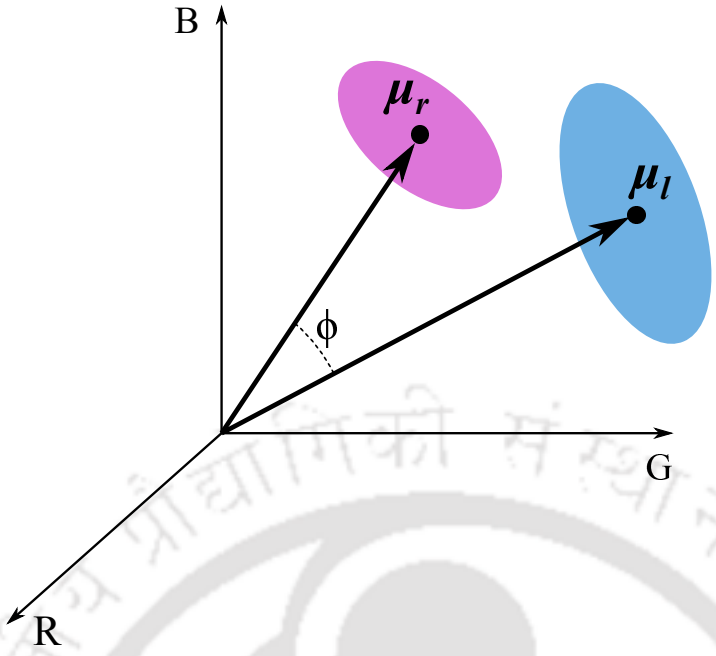


Figure 4.2: Similarity between two clusters.

and the overlap between G_i and G_j is given by:

$$\varepsilon_{i,j} = \exp\left[-d_{Bh}\left(G_i, G_j\right)\right] \ \forall i, j \tag{4.5}$$

In this paper, the standard Bhattacharyya distance is termed as Original Bhattacharyya Distance (OBD).

Now, let us consider two Gaussian clusters of pixels in RGB space – a reference cluster $C_r\left(\mu_r, \mathbf{\Sigma}_r\right)$ and its modified version $C_l\left(\mu_l, \mathbf{\Sigma}_l\right)$ on account of local illumination change. Therefore, distance between C_l and C_r should be ideally zero as both of them correspond to similar image regions. However, OBD between C_l and C_r is a non-zero value. The OBD treats the two distribution functions as two different independent entities. However, two distribution functions, which are actually originated from similar regions (*e.g.*, skin regions with colour deformations) have some correlations, and that is why, the distance between them should be less than the OBD value. To overcome this problem, the OBD is modified in such a way that $d_{MBh}\left(G_l, G_r\right) < d_{Bh}\left(G_l, G_r\right)$ for the distribution functions corresponding to similar image regions. We termed the proposed distance measure $d_{MBh}\left(G_l, G_r\right)$ as Modified Bhattacharyya Distance (MBD). Local variations of skin colours are approximated as a combination of two independent parameters– orientation of the centroid vector μ_l with respect to μ_r, and the ori-

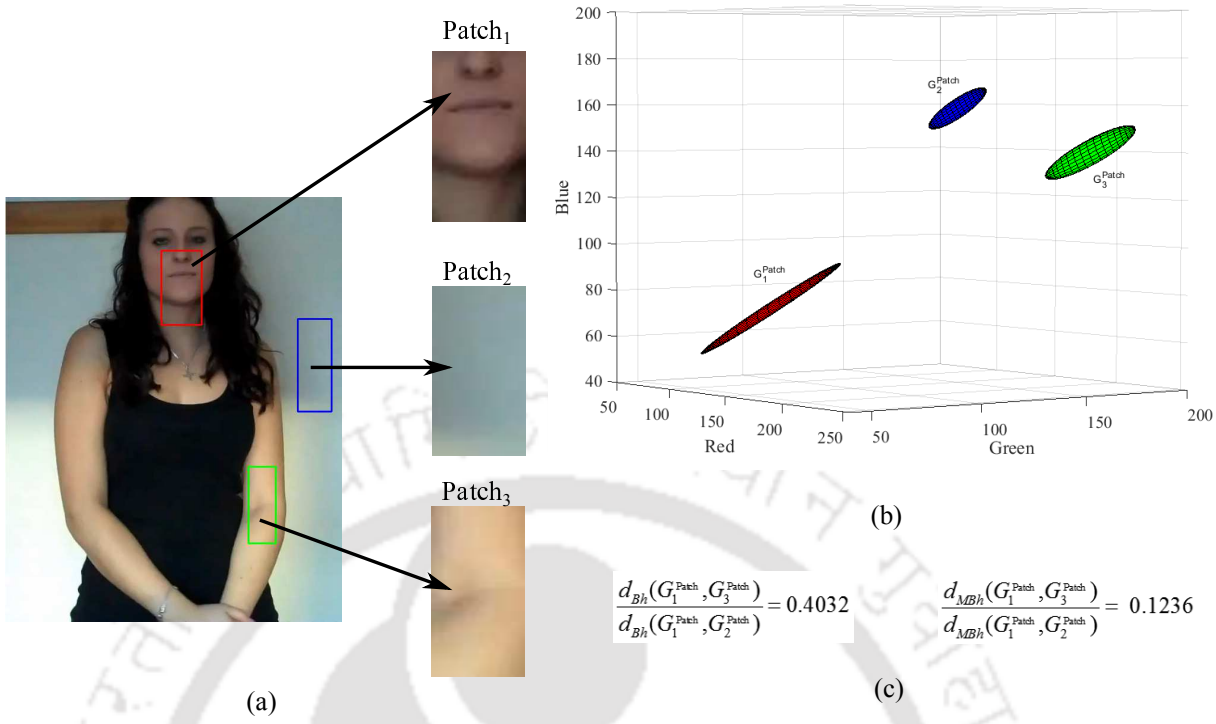


Figure 4.3: Effect of using MBD as distance measure – (a) original image with three patches extracted from different regions; (b) distribution of pixels in RGB space for different patches – red cluster corresponds to Patch_1, blue cluster corresponds to Patch_2, green cluster corresponds to Patch_3; and (c) distance ratios for different patches.

entation of C_l with respect to C_r as shown in Fig. 4.2. Let, ϕ be the angle between μ_l and μ_r. The two clusters will be perfectly aligned if $\left[1 - \frac{tr(\mathbf{V}_l^T \mathbf{V}_r)}{3}\right] = 0$, where, $\mathbf{V}_l$ and $\mathbf{V}_r$ are the eigen vector matrices of $\mathbf{\Sigma}_l$ and $\mathbf{\Sigma}_r$ respectively. On the other hand, smaller is the value of ϕ, the more chromatically similar are the clusters. Hence, the closed form of MBD is expressed as:

$$\begin{aligned} d_{MBh}(G_i, G_j) =& \xi(\mu_i - \mu_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2}\right]^{-1} (\mu_i - \mu_j) \\ &+ \gamma \ln \frac{|(\Sigma_i + \Sigma_j)/2|}{\sqrt{|\Sigma_i| \cdot |\Sigma_j|}}, \ \forall i, j \end{aligned} \tag{4.6}$$

where,

$$\begin{aligned} \xi = \tfrac{1}{8}\left[1 - \exp\left(-\tfrac{\phi}{\phi_{\max}}\right)\right], \gamma = \tfrac{1}{2}\left[1 - \tfrac{tr(\mathbf{V}_i^T \mathbf{V}_j)}{3}\right] \\ \phi = \cos^{-1}\left(\tfrac{\langle \mu_i, \mu_j \rangle}{\|\mu_i\| \cdot \|\mu_j\|}\right) \end{aligned} \tag{4.7}$$

Here, ϕ_{max} is the maximum allowed deviation in centroid orientation. Following rule is framed

on the basis of the equation 4.6.

$$d_{MBh} \rightarrow d_{Bh} : (\phi \gg \phi_{\max}) \wedge \left\{ \frac{tr\left(\mathbf{V}_i^T \mathbf{V}_j\right)}{3} \ll 1 \right\} \equiv true \tag{4.8}$$

Fig. 4.3 illustrates the effect of using MBD over OBD in our proposed method. As shown in Fig. 4.3-a, three patches are extracted from an image, they are $Patch_1$, $Patch_2$, and $Patch_3$. The $Patch_1$ and $Patch_3$ belong to skin regions, whereas $Patch_2$ belongs to non-skin regions. We consider $Patch_1$ has actual skin tone and $Patch_3$ has a deformed skin tone due to local illumination variation. Fig. 4.3-b shows the distribution of pixels belonging to these patches in RGB space. In Fig. 4.3-c, distance ratios calculated by using OBD and MBD are shown for these image patches. From this analysis, it is evident that the relative distance between an unknown skin distribution and a reference skin distribution is less for the MBD in contrast to the OBD.

Now, the overlap between Gaussian components of G^I and FSDM is then given by:

$$\varepsilon_i^f = \exp\left[-d_{MBh}\left(G_i^I, G^f\right)\right] \text{ for } i = 1, .., K_I \tag{4.9}$$

and the overlap between Gaussian components of G^I and G^{gb} is given by:

$$\varepsilon_i^{gb} = \exp\left[-\min_{\forall j} d_{MBh}\left(G_i^I, G_j^{gb}\right)\right] \text{ for } j = 1, .., K_{gb} \tag{4.10}$$

A Gaussian component in G^I should belong to background regions if it overlaps less with the FSDM than any of its overlapping with the GBDM components. Also, background regions may be chromatically similar to the skin regions in some of the cases. Therefore, the local background distribution model (LBDM) G^{lb} should include only those Gaussian components which follow inclusion criterion as defined below:

$$G_i^I \in \left\{ G^{lb} : \left(\varepsilon_i^f \leq \varepsilon_i^{gb}\right) \vee \left(\varepsilon_i^f \leq \tau_1\right) \equiv true \right\} \tag{4.11}$$

The weights of the Gaussian components in G^{lb} are derived from the weights they had in G^I followed by a normalisation. It is found that the GBDM is better in discriminating skin-like colours from other colours, whereas the LBDM better discriminates actual skin colours from skin-like colours in the background. Hence, a fusion of GBDM and LBDM can provide better

discrimination between actual skin pixels and non-skin pixels as compared to either GBDM or LBDM. Accordingly, the fusion-based background distribution model (FBDM) *i.e.*, G^{fb} is expressed as:

$$G^{fb}(\mathbf{X}) = max\left\{G^{lb}(\mathbf{X}), G^{gb}(\mathbf{X})\right\} \tag{4.12}$$

where, $\mathbf{X} = [R\,G\,B]'$ is the colour vector of a pixel.

4.2.2 Dynamic module

As explained earlier, the moving body parts cause local illumination variations in the scene even if the global illumination is kept constant. This non-uniform illumination may change the chromatic appearance of moving skin regions in such a manner that they may become undetectable by the static module. Therefore, the skin model should adapt to local variations of skin colours due to non-uniform illumination. However, updating a skin model for every frame is computationally expensive for real-time applications. Therefore, the skin model should be updated only when a significant change in scene colour takes place due to local illumination variations, and we proposed a key frame detection technique to update the skin model only for the key frames.

4.2.2.1 Key frame selection

The key frames in a video sequence are selected on the basis of the measure of change in chromaticity — for a given key frame, the next key frame is the one in which chromaticity changes significantly. By doing so, an entire video clip is transformed into a small number of representative key frames. The chromatic changes between a frame and a reference frame can be determined from chromatic entropies of both the frames. Let, I_t and I_{ref} represent the frame at t and the reference frame, respectively. The change in chromatic entropy in I_t with respect to I_{ref} is given by:

$$\Delta E_t = \frac{|E_t - E_{ref}|}{E_{ref}} \tag{4.13}$$

where,

$$E_t = -\frac{\sum\limits_{\forall \mathbf{X} \in I_t} n_{\mathbf{X}} \log\left(\frac{n_{\mathbf{X}}}{w \times h}\right)}{w \times h} \tag{4.14}$$

Here, E_t and E_{ref} represent chromatic entropy of I_t and I_{ref}, respectively; $n_{\mathbf{X}}$ is the number of occurrence of the colour vector $\mathbf{X} = [RGB]'$ in I_t; w is the frame width and h is the frame height. A frame I_t is designated as a key frame I_{key} if its ΔE_t is greater than some threshold value θ_E.

4.2.2.2 Moving skin distribution model

Assuming the background is static, the frame difference method is the most easiest way to detect moving objects in a frame. However, the frame difference algorithm suffers from two major limitations – occurrence of ghost foreground regions and foreground apertures as shown in Fig. 4.4. Ghost foreground regions occur due to the motion of the objects. During frame differencing, an ambiguity may occur between real foreground regions and ghost foreground regions. The other drawback of frame difference is the occurrence of foreground object aperture (FOA). FOA occurs if the object is texture-less and/or the intensity gradient of the image is significantly less. The probability of occurrence of FOA is significantly high in case of moving skin regions as they have less texture and intensity gradient. In order to avoid the occurrence of ghost foreground regions, Kameda and Michihiko [127] proposed a "double-difference of frame (DDF)" method. In this method, three frames at time $t-2$, $t-1$, and t are selected. The DDF method performs a logical AND operation over thresholded difference frames between frames at $t-2$ and $t-1$, and frames at $t-1$ and t. The result of a DDF algorithm in presence of FOA is shown in Fig. 4.5. In this, O_{t-2}, O_{t-1}, O_t show object positions at time instants $t-2$, $t-1$, and t, respectively.

The DDF algorithm produces a narrow region for a moving object if the object has less texture and/or intensity gradient. An use of morphological operations can reduce FOA in a difference frame as shown in Fig. 4.6. A dilation along the direction of motion of an object can reduce FOA with an inclusion of ghost foreground regions. Motivated by this fact, a morphological enhancement-based double difference frame method is proposed to detect moving skin regions. In our proposed method, morphological dilation is applied on each of the thresholded difference frames. Subsequently, a logical OR operation is performed over the dilated difference frames. The OR operation helps in including more moving skin regions between the consecutive frames. However, this operation significantly increases occurrence of ghost foreground regions. In order to reduce inclusion of background regions in foreground regions of a thresholded differ-

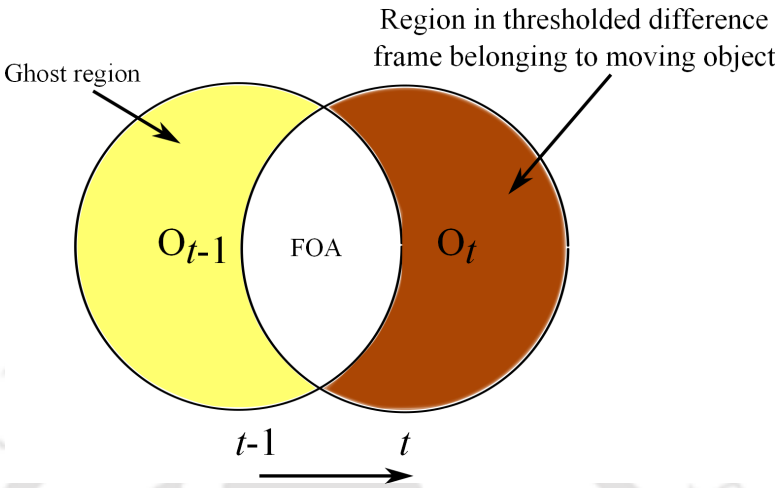


Figure 4.4: Drawback of single frame difference method.

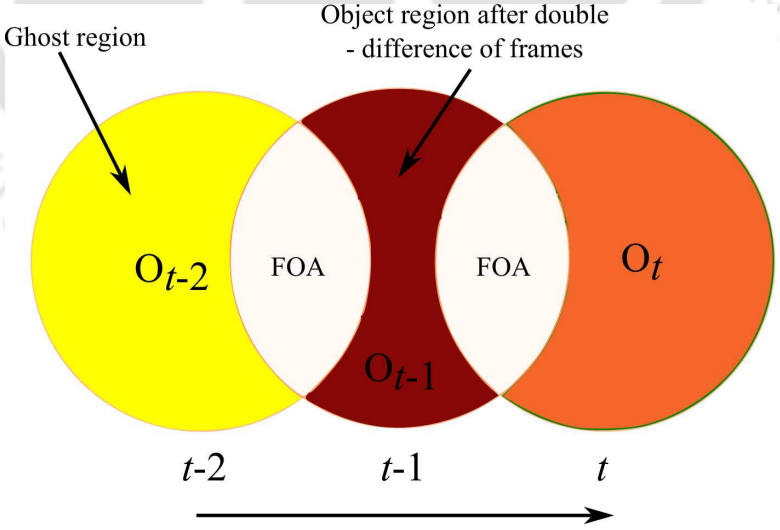


Figure 4.5: Results of DDF method in presence of FOA

ence frame, dilation should be performed in the direction of motion of the foreground objects as shown in Fig.4.6. However, in case of articulated objects like hands, foreground motion could be complex. We approximate complex movements as a combination of motions in four directions - $0°, 45°, 90°, 135°$ with respect to the horizontal direction. Directional opening can be used to select a region in a particular direction. After directional opening, a dilation in the perpendicular direction of the opening process grows a region in the direction of its motion as shown in Fig. 4.7. Finally, a logical OR operation is performed on the two morphologically enhanced thresholded difference frames to obtain a moving object mask $BW_{\Delta I}^{final}$. Fig. 4.8 shows the flowchart of our proposed morphology-based moving object localization method. An example of the proposed moving skin region localization approach is shown in Fig. 4.9.

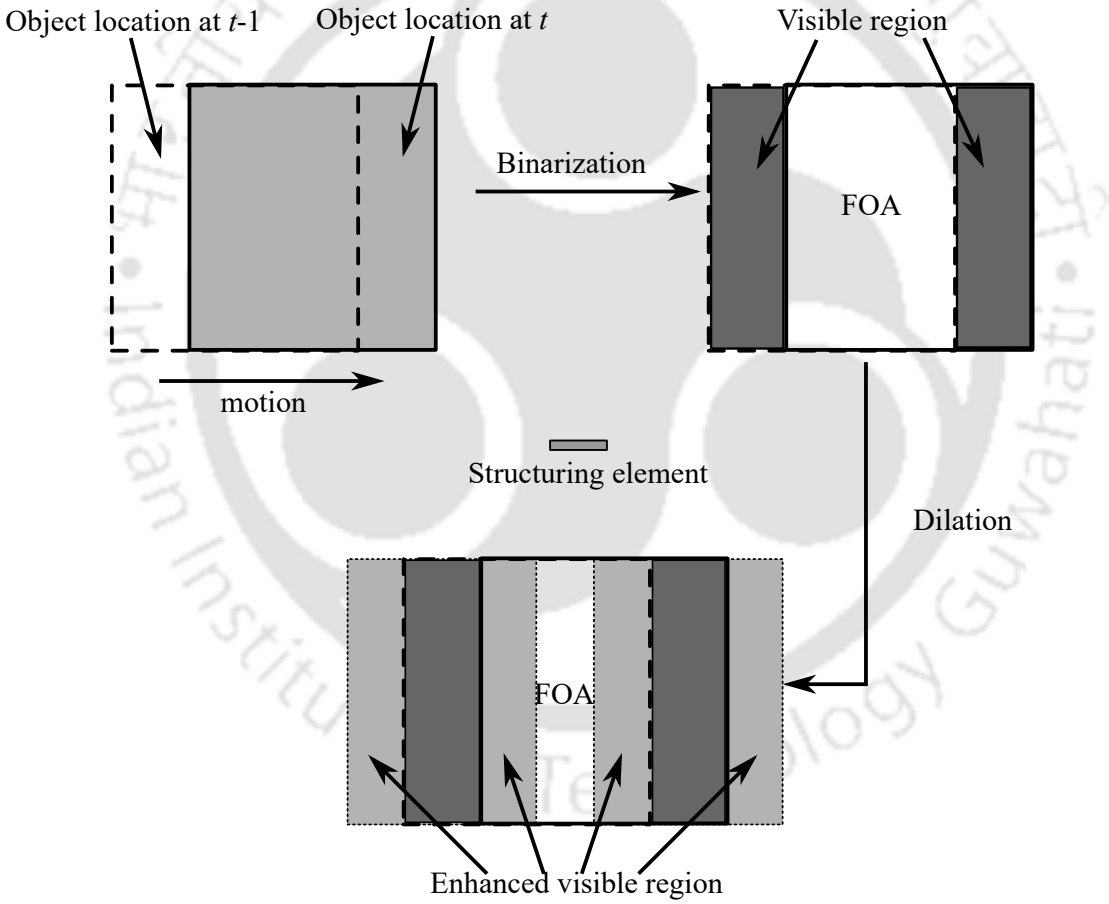


Figure 4.6: Reduction in FOA using directional dilation.

The moving object mask $BW_{\Delta I}^{final}$ is used for extracting moving skin regions at $t-1$. As the chromatic changes of the skin coloured regions are temporally correlated, the colour of skin regions in I_t can be approximated by the colour of skin regions in I_{t-1}. The mask $BW_{\Delta I}^{final}$ is applied on I_{t-1} to extract a moving object pixel set Q_{t-1}. The current frame I_t and its

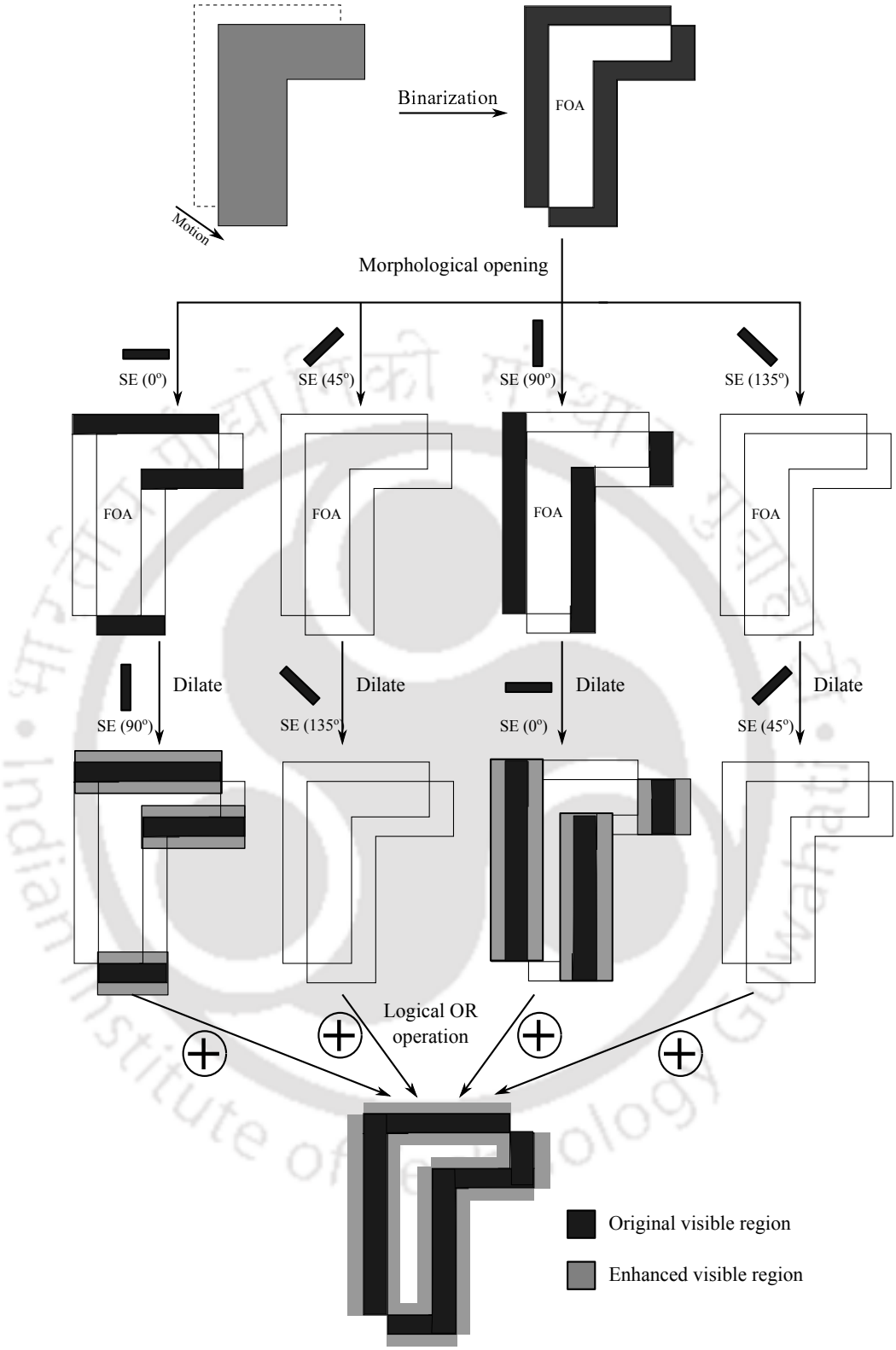


Figure 4.7: Proposed method for reducing FOA using directional "opening" followed by directional "dilation".

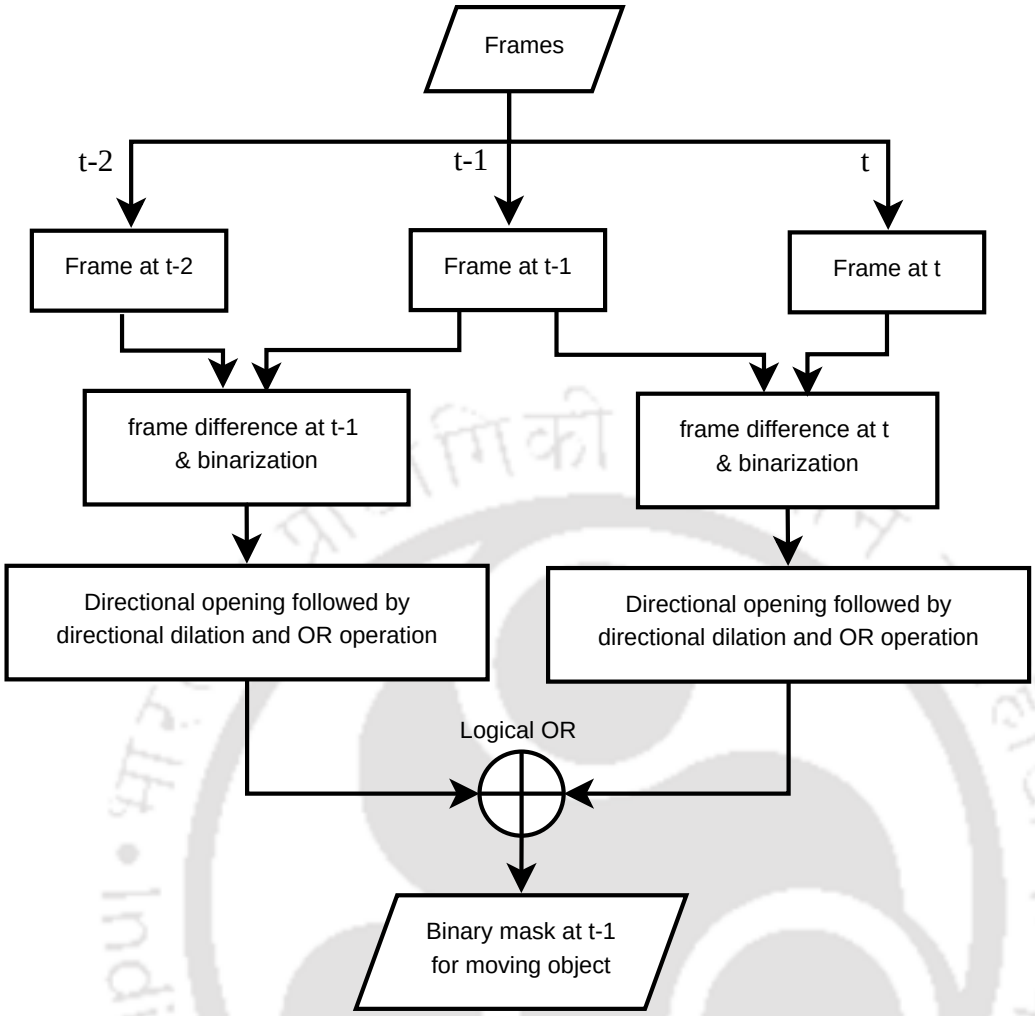


Figure 4.8: Flowchart of the modified DDF algorithm

last N_{TF} number of previous frames are considered to derive Q_ts for last N_{TF} frames. Finally, all these Q_ts are stacked hierarchically in a new pixel set Q_{motion} with descending values of t. Algorithm 2 is the proposed algorithm to derive the moving object pixel set Q_{motion}. Here, N_f is the total number of frames in a video. Ideally Q_{motion} should only contain pixels belonging to moving skin regions. So, the pixel set Q_{motion} can be used to derive a colour distribution model for moving skin regions, *i.e.*, the moving skin distribution model (MSDM). However, some of the pixels from non-skin foreground (*e.g.* clothing) and/or background regions may be included in Q_{motion} due to the presence of ghost regions and the proposed morphological region enhancement. Therefore, a filtering process is necessary to exclude this background pixels to derive the MSDM. To obtain the MSDM, a GMM is used to model the distribution of all the

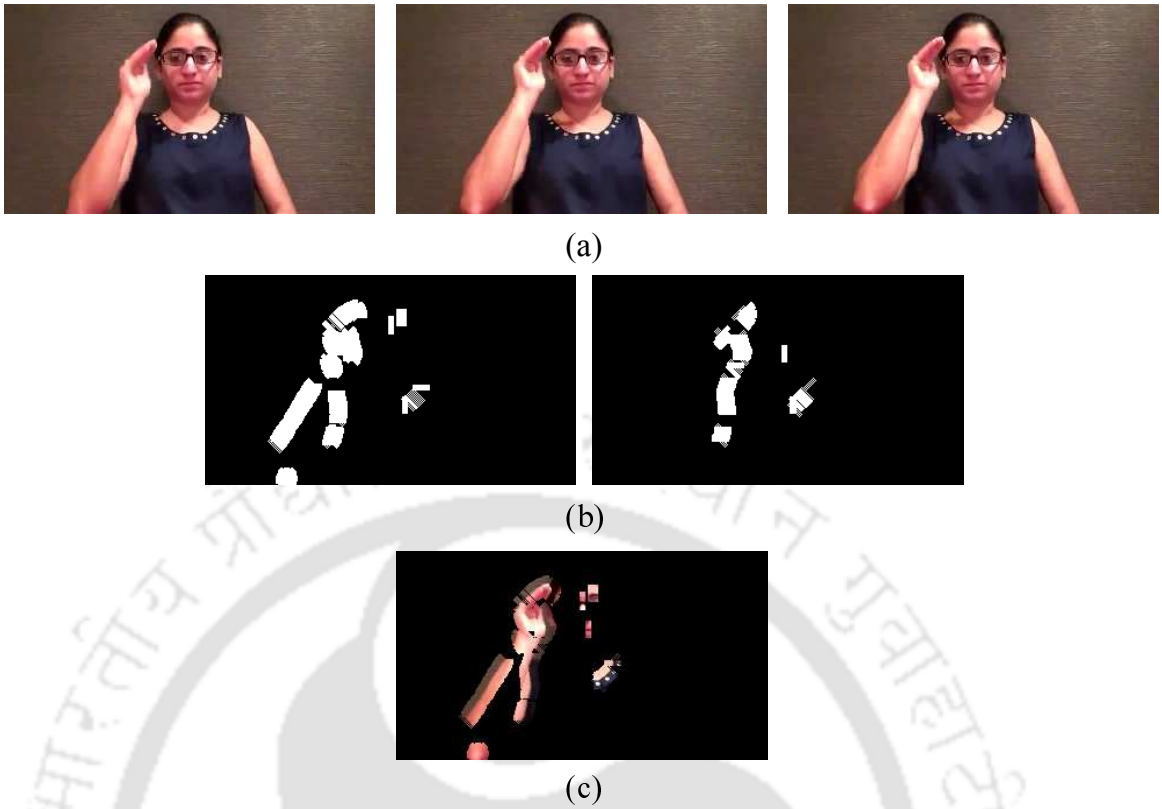


Figure 4.9: Example of morphology-based moving object localization: (a) frames at $t-2, t-1$ and t, (b) morphologically enhanced binarized difference frames at $t-1$ and t, and (c) final localized moving object regions.

pixels in Q_{motion} as:

$$G_m^{init} = \sum_{k=1}^{K_m^{init}} \omega_m^{init} N\left(\mu_k^{init}, \Sigma_k^{init}\right) \tag{4.15}$$

where, G_m^{init} represents an initial pixel distribution model for moving objects. Some of the Gaussian components in G_m^{init} correspond to background and/or non-skin moving object regions. These Gaussian components should be nearer to the background model FBDM as compared to FSDM. The proposed MBD measure is used to determine the overlapping of G_m^{init} components with FSDM and FBDM. The overlap between Gaussian components of G_m^{init} and G^f is given by:

$$\varepsilon_i^{f,m} = \exp\left[-d_{MBh}\left(G_{m,i}^{init}, G^f\right)\right] \text{ for } i = 1, .., K_m^{init}, \tag{4.16}$$

Similarly, the overlap between Gaussian components of G_m^{init} and the global background model

Algorithm 2: Proposed algorithm to derive the moving object pixel set Q_{motion}

Data: I_t where, $t = 1, 2, .., N_f$
Result: Moving object pixel set Q_{motion}
for $N_{TF} < t \leq N_f$ **do**
  if $t = N_{TF} + 1$ **then**
    for $3 \leq k \leq t$ **do**
      Obtain:
        $\Delta I_{k-1} \leftarrow |I_{k-1} - I_{k-2}|$
        $\Delta I_k \leftarrow |I_k - I_{k-1}|$
        $BW_{\Delta I_j} \xleftarrow{binirization} \Delta I_j$ where, $j = k, k-1$.
      Perform: Morphological region enhancement on $BW_{\Delta I_k}$ and $BW_{\Delta I_{k-1}}$.
      Obtain:
        $BW_{\Delta I}^{final} \leftarrow BW_{\Delta I_{k-1}} \cup BW_{\Delta I_k}$
      Obtain:
        $Q_{k-1} \leftarrow I_{k-1}(x) : BW_{\Delta I}^{final}(x) = 1$
      Add Q_{k-1} to the bottom of Q_{motion}.
    end
    Perform:
      $I_{key} \leftarrow I_t$
  else
    Calculate ΔE_t
    if $\Delta E_t > \theta_E$ **then**
      Remove $Q_{t-N_{TF}}$ from the top of Q_{motion}.
      Obtain: Q_t
      Add Q_t to the bottom of Q_{motion}.
      Update:
        $I_{key} \leftarrow I_t$
    else
      Do not update I_{key}
    end
  end
end
return Q_{motion}

G^{gb} is given by:

$$\varepsilon_i^{gb,m} = \exp\left[-\min_{\forall j} d_{MBh}\left(G_{m,i}^{init}, G_j^{gb}\right)\right] \text{ for } j = 1, .., K_{gb}, \tag{4.17}$$

and the overlap between Gaussian components of G_m^{init} and the local background model G^{lb} is given by:

$$\varepsilon_i^{lb,m} = \exp\left[-\min_{\forall j} d_{MBh}\left(G_{m,i}^{init}, G_j^{lb}\right)\right] \text{ for } l = 1, .., K_{lb} \tag{4.18}$$

The overall overlapping of G_m^{init} with FBDM is expressed as:

$$\varepsilon_i^{fb,m} = max\left(\varepsilon_i^{gb,m}, \varepsilon_i^{lb,m}\right), \forall i = 1, .., K_m^{init} \tag{4.19}$$

Finally, an inclusion criterion is proposed for Gaussian components of G_m^{init} to the final moving skin distribution model (MSDM) *i.e.*, G_m^{skin}. The inclusion criterion is formulated as follows:

$$G_{m,i}^{init} \in \left\{ G_m^{skin} : \left(\varepsilon_i^{f,m} > \varepsilon_i^{fb,m} \right) \wedge \left(\varepsilon_i^{f,m} > \tau_2 \right) \equiv true \right\} \tag{4.20}$$

4.2.3 Derivation of a skin mask

[1] The pixels of each of the frames are classified into skin and non-skin pixels by using a Bayesian classifier [13]. The Bayesian classifier provides a Skin Probability Map (SPM) for a frame at t. The SPM at a location x represents posteriori probability of a pixel $\mathbf{X}$ belonging to skin region at that location, and it is defined as:

$$SPM(x) = \frac{P(S) \cdot P(\mathbf{X}\,|S)}{P(S) \cdot P(\mathbf{X}\,|S) + P(NS) \cdot P(\mathbf{X}\,|NS)} \tag{4.21}$$

where, $P(\mathbf{X}\,|S)$ and $P(\mathbf{X}\,|NS)$ are likelihoods, $P(S)$ and $P(NS)$ are priors for skin (S) and non-skin (NS) pixels, respectively. Hence, the SPM derived using FSDM and FBDM is given by:

$$SPM_f(x_t) = \frac{P(S)G^f(\mathbf{X}_t)}{P(S)G^f(\mathbf{X}_t) + P(NS)G^{fb}(\mathbf{X}_t)} \tag{4.22}$$

The SPM value derived using MSDM and FBDM is given by:

$$SPM_m(x_t) = \frac{P(S) \cdot G_m^{skin}(\mathbf{X}_t)}{P(S) \cdot G_m^{skin}(\mathbf{X}_t) + P(NS) \cdot G^{fb}(\mathbf{X}_t)} \tag{4.23}$$

It is observed that SPM_f can capture static skin regions more accurately, whereas SPM_m is more responsive to moving skin regions. Therefore, the final SPM at time instant t is expressed as:

$$SPM_{final}(x_t) = \max\{SPM_f(x_t), SPM_m(x_t)\} \tag{4.24}$$

The final skin mask $Mask$ for a frame at t is obtained by thresholding SPM_{final} with an appropriate threshold θ_{th} as:

$$Mask(x_t) = \begin{cases} 1 \text{ if } SPM_{final}(x_t) \geq \theta_{th} \\ 0 \text{ otherwise.} \end{cases} \tag{4.25}$$

[1] This work is in under review in IEEE Trans. Circuits and Systems for Video Technology.

4.3 Experimental Analysis

4.3.1 Experimental setup

In order to obtain a global background model G^{gb}, background samples are extracted from a set of randomly selected images of ECU dataset [65]. For experimental validation of our proposed algorithm, a set of sign language videos collected from the web are used. The selected videos are captured in unconstrained illumination and different background conditions. Each of these videos has a duration of 10 seconds with varying frame rate, spatial resolution and illumination conditions. For quantitative performance analysis, these videos are manually annotated for skin and non-skin regions. From the annotated videos, it is observed that $P(S) \sim [0.1, 0.2]$ for standard definition (SD) videos (4:3 aspect ratio), and $P(S) \sim [0.05, 0.15]$ for high definition (HD) videos (16:9 aspect ratio). Hence, we judiciously choose $P(S) = 0.15$ for SD videos and $P(S) = 0.1$ for HD videos. The respective values for $P(NS)$ are derived as, $P(NS) = 1 - P(S)$. Four quantitative measures namely, *detection accuracy (Acc.)*, *false positive error rate* (δ_{fp}), *false negative error rate* (δ_{fn}), *total detection error rate*(δ_t) are selected for evaluation, where $\delta_t = \delta_{fp} + \delta_{fp}$. All the detection errors are obtained by thresholding SPM_{final} with a threshold θ_{th}. For simplicity of implementation, we select $\tau_1 = \tau_2 = \tau$.

4.3.1.1 Determination of τ

The parameter τ controls the inclusion of Gaussian components in G^{lb} and G_m^{skin}. A smaller value of τ results in inclusion of Gaussian components into G_m^{skin} which belong to skin-like background colours. This increases the chance of false acceptance. On the other hand, a larger value of τ results in inclusion of some of the Gaussian components into G^{lb}. This increases the chance of false rejection. Hence, we have proposed a selection scheme for the parameter τ as follows:

Let, I_t be the frame at t. The chromatic randomness of an image can be determined by using its chromatic entropy. In our previous work [23], it is shown that the chromatic entropy of a skin-masked image, also termed as skin-chroma entropy (E_s) can be used as a measure for

probable false acceptance error. The parameter E_s can be obtained as:

$$E_s = -\frac{\sum\limits_{\forall y \in \mathbf{Y}} n_{\mathbf{X}(y)} \log\left(\frac{n_{\mathbf{X}(y)}}{N_{\mathbf{Y}}}\right)}{N_{\mathbf{Y}} \cdot \log(255^3)} \tag{4.26}$$

where,

$$y \in \left\{\mathbf{Y} : P_{rgb}^{skin}(y) \geq \theta_{sp}\right\} \quad , \tag{4.27}$$

and

$$P_{rgb}^{skin}(y) = \frac{P(S) \cdot G^f(\mathbf{X})}{P(S) \cdot G^f(\mathbf{X}) + P(NS) \cdot G^{gb}(\mathbf{X})} \tag{4.28}$$

Here, $n_{\mathbf{X}(y)}$ is the number of count of colour vector $\mathbf{X}(\mathbf{y})$ and $N_{\mathbf{Y}}$ is the total number of pixel locations in $\mathbf{Y}$. The threshold θ_{sp} is obtained by using Otsu's method [116]. Similarly, a face-masked version of I_t *i.e.*, I_{face} is obtained by applying a face mask derived by using the Viola and Jones algorithm [125]. Let, E_f be the chromatic entropy of I_{face}. Ideally, E_s should be equal to E_f. However, in practice I_{skin} may contain skin-like coloured background regions along with some of the true skin regions. So, $E_s > E_f$ in presence of skin-like background colours. Thus, we express the parameter τ as:

$$\tau = \left[\sum_n \left(\frac{E_f}{E_s}\right)^{n/2}\right]^{-1} \simeq \left[1 + \sqrt{\frac{E_f}{E_s}} + \frac{E_f}{E_s}\right]^{-1} \tag{4.29}$$

where, $n = 0, 1, 2, ...$ As the ratio $\frac{E_f}{E_s} < 1$, higher order powers $(n > 2)$ are neglected.

Table 4.1: Detection results for different values of ϕ_{max}

ϕ_{max}	$\delta_{fp,avg}$ (%)	$\delta_{fn,avg}$ (%)	$\delta_{t,avg}$ (%)	**Avg. Acc.**
0°	6.41	14.34	20.75	0.9250
10°	6.04	9.48	15.52	0.9362
20°	**5.78**	**8.19**	**13.97**	**0.9404**
30°	7.00	8.29	15.29	0.9301
40°	11.09	8.53	19.62	0.8959
50°	11.52	7.92	19.44	0.8934

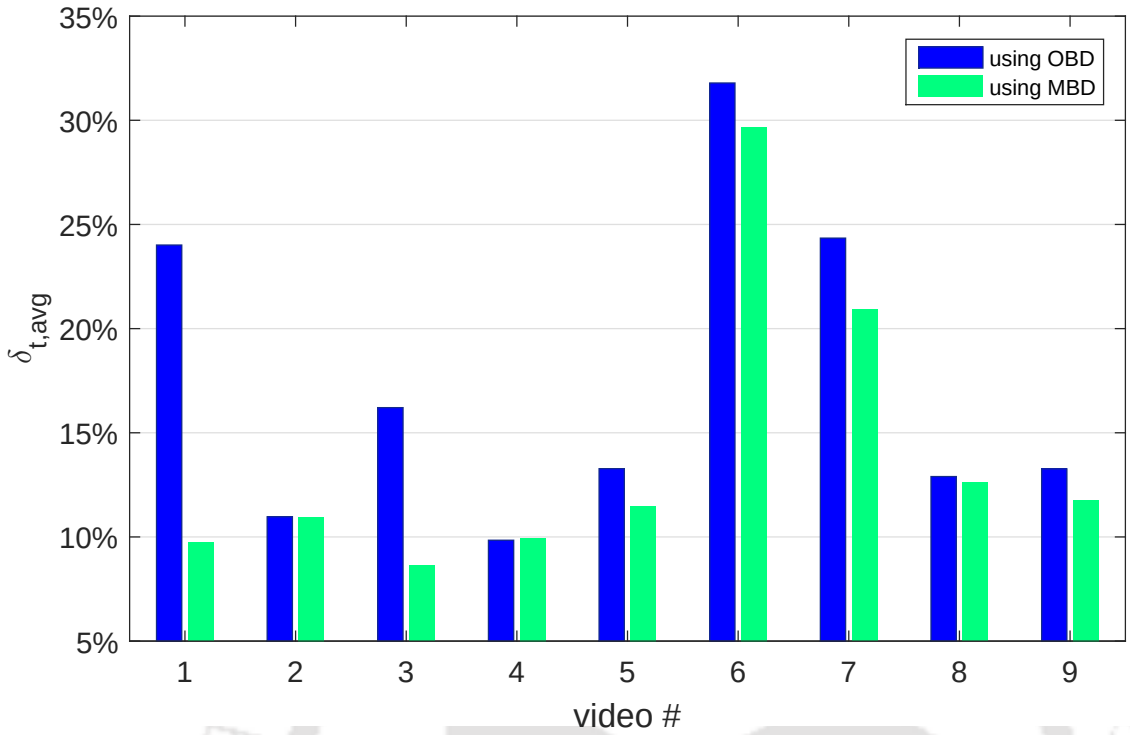


Figure 4.10: Video-wise comparison between OBD and MBD.

Table 4.2: Comparative analysis for OBD and MBD

Method	$\delta_{fp,avg}$ (%)	$\delta_{fn,avg}$ (%)	$\delta_{t,avg}$ (%)	Avg. Acc.
SPM_{final} using **OBD**	6.13	11.28	17.41	0.9326
SPM_{final} using **MBD**	**5.78**	**8.19**	**13.97**	**0.9404**

4.3.2 Experimental validation

At first, we examine the effect of ϕ_{max} on detection results. For this, detection errors are calculated by varying ϕ_{max}, and the results are given in Table 4.1. It is observed that when $\phi_{max} = 0°$, average false negative error $\delta_{fn,avg}$ is maximum. For $\phi_{max} = 0°$, the parameter $\xi = \frac{1}{8}$. This implies that some of the clusters corresponding to the true skin pixels in G^I are included in LBDM, and/or excluded from MSDM. In either of these conditions, clusters whose centroids are very close to that of FSDM are only selected as skin clusters. An increase in ϕ_{max} value makes more skin-like clusters to be excluded from the background model, and/or included in MSDM. Thus, false negative error reduces with an increase in ϕ_{max} value. However, this increases the chance of false acceptance error. Some of the clusters of skin-like background

Table 4.3: Comparative analysis of different methods

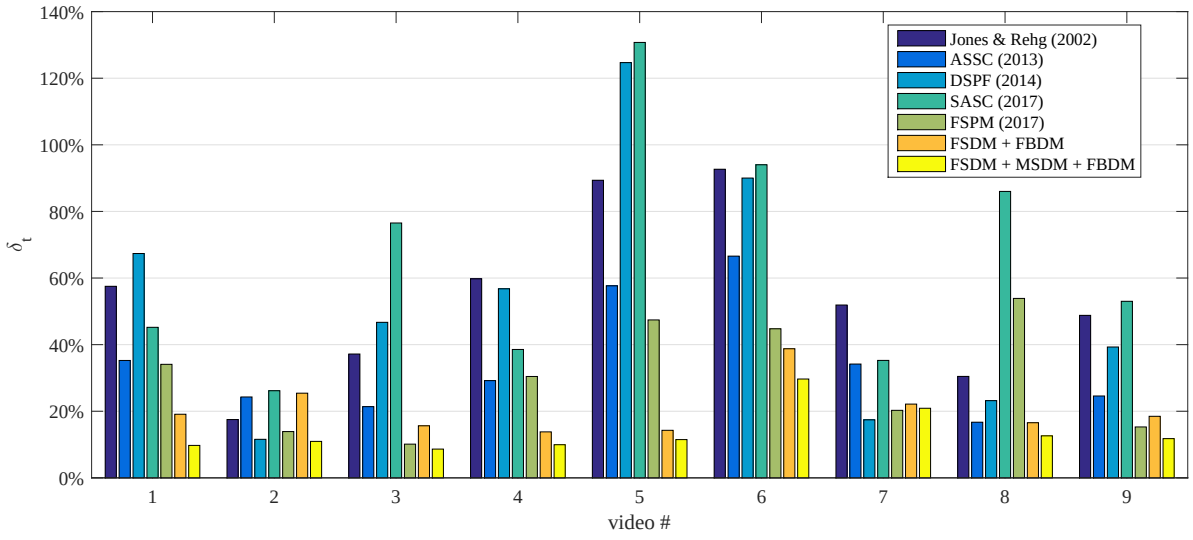
Method	$\delta_{fp,avg}$ (%)	$\delta_{fn,avg}$ (%)	$\delta_{t,avg}$ (%)	Avg. Acc.
Bayes classifier [13]	19.69	34.22	53.91	0.7842
SDDMA [128]	4.71	58.08	62.80	0.8710
ASSC [21]	8.57	25.85	34.42	0.8919
DSPF [9]	21.67	22.37	44.04	0.7832
SASC [22]	11.70	45.15	56.84	0.8470
FSPM [23]	16.85	13.17	30.02	0.8373
FSDM + FBDM	**5.33**	15.14	20.47	0.9345
FSDM + MSDM + FBDM	5.78	**8.19**	**13.97**	**0.9404**

δ_{fp} = false positive error rate; δ_{fn} = false negative error rate; δ_t = total detection error rate = $\delta_{fp} + \delta_{fn}$; *Acc.* = *Accuracy*

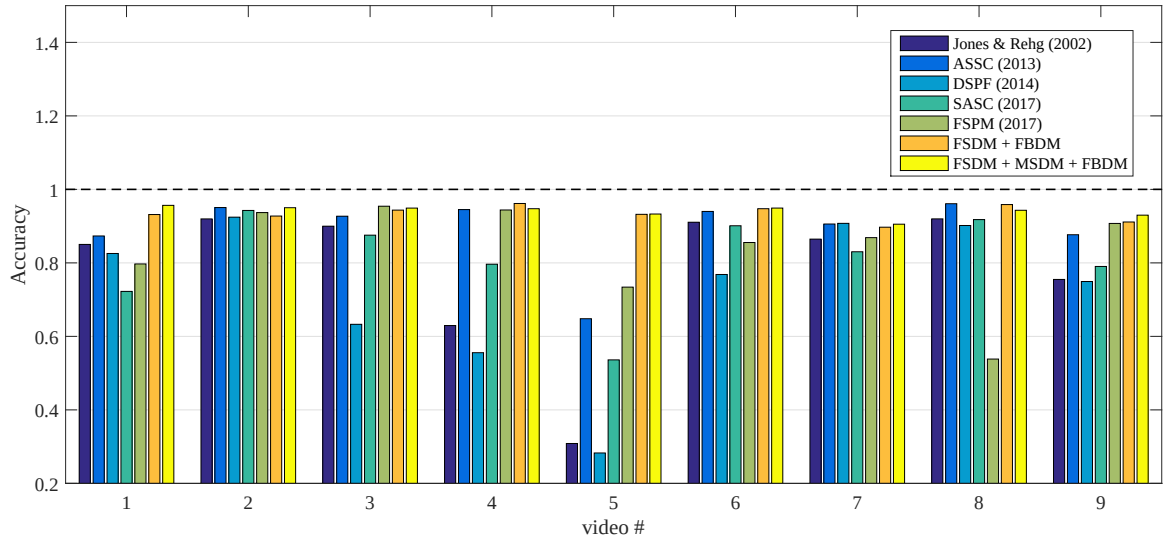
pixels can be included in MSDM and/or excluded from LBDM if the ϕ_{max} is increased further. For example, the average false positive error $\delta_{fp,avg}$ increases significantly for an increase in ϕ_{max} from 20° to 40°. On the other hand, the average false negative error $\delta_{fn,avg}$ decreases significantly for an increase in ϕ_{max} from 0° to 20°. It is also observed that for $\phi_{max} \geq 20°$, the average total detection error $\delta_{t,avg}$ for all the videos becomes the lowest. Hence, ϕ_{max} value is fixed at 20° for the remaining analysis.

A comparative analysis is also performed between the original Bhattacharyya distance (OBD) and the proposed Modified Bhattacharyya distance (MBD). Detection errors are calculated by both OBD and MBD. As mentioned in Eq. 4.8, the maximum value of MBD is fixed by OBD. The comparison between the OBD and the MBD in calculating SPM is Shown in Table 4.2. In contrast to OBD, detection errors reduce due the application of MBD. This validates the efficacy of the proposed MBD in skin detection.

Finally, the proposed method is compared with the state-of-the-art methods, such as – Bayes classifier [13], Fast propagation-based skin segmentation (FPSS) [14], Adaptive seed-based skin classification (ASSC) [21], Skin detection by dual maximization of detectors agreement (SDDMA) [128], Discriminative Skin-Presence Feature (DSPF) [9], Stacked Autoencoders-based skin classification (SASC) [22], and Fusion-based Skin Probability Map (FSPM) [23]. The detection results obatined by different methods are given in Table 4.3. The Bayes classifier method



(a)



(b)

Figure 4.11: Comparative bar plots for different videos: (a) δ_t for different videos, and (b) Accuracy for different videos.

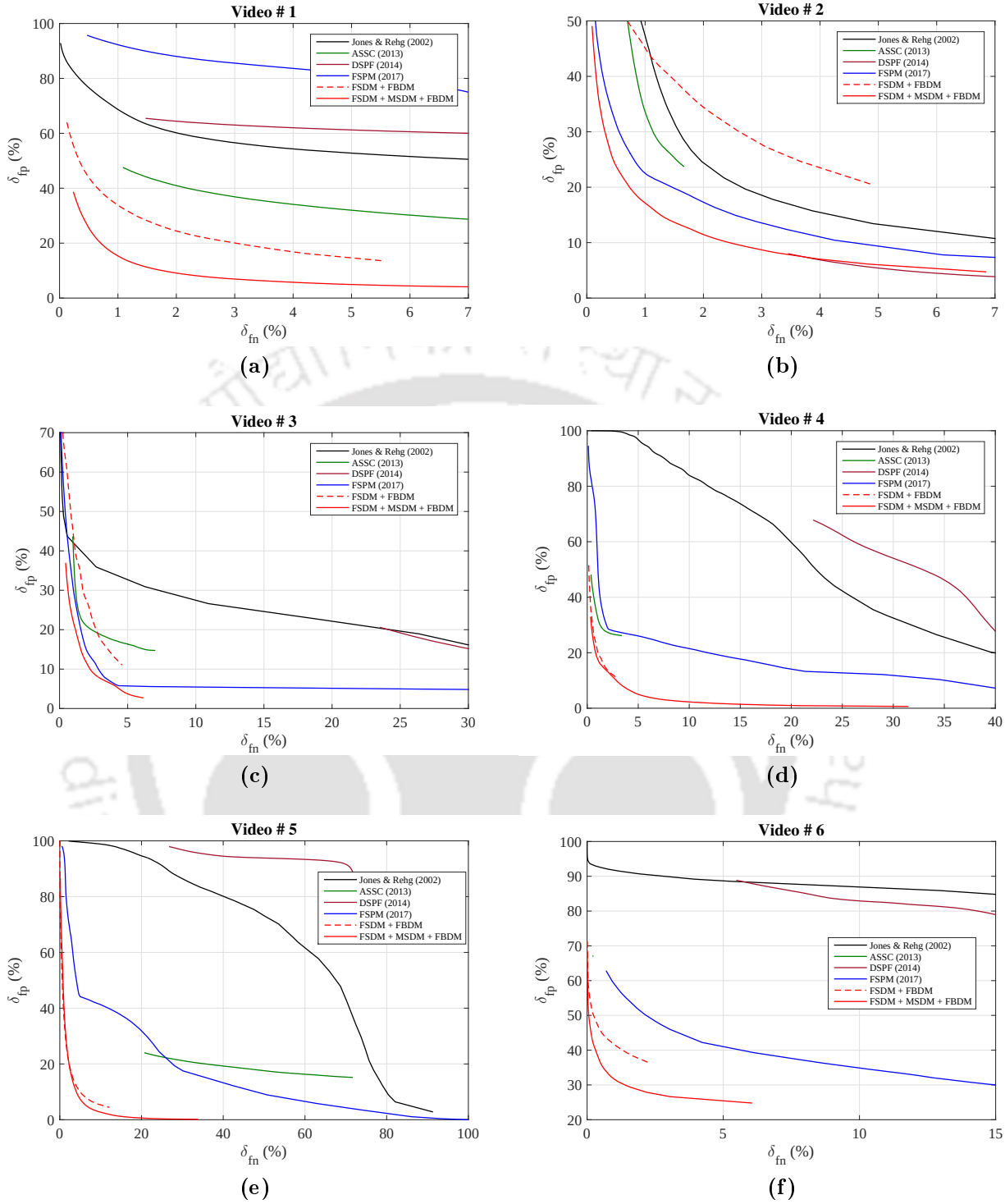


Figure 4.12: ROC plots for different videos (1–6).

proposed by Jones and Rehg [13] is considered as a benchmark. For training, the method needs a set of skin and non-skin pixel samples, which are obtained globally. However, the accuracy of this method is totally dependent on the training sample set. The Jones & Rehg's method

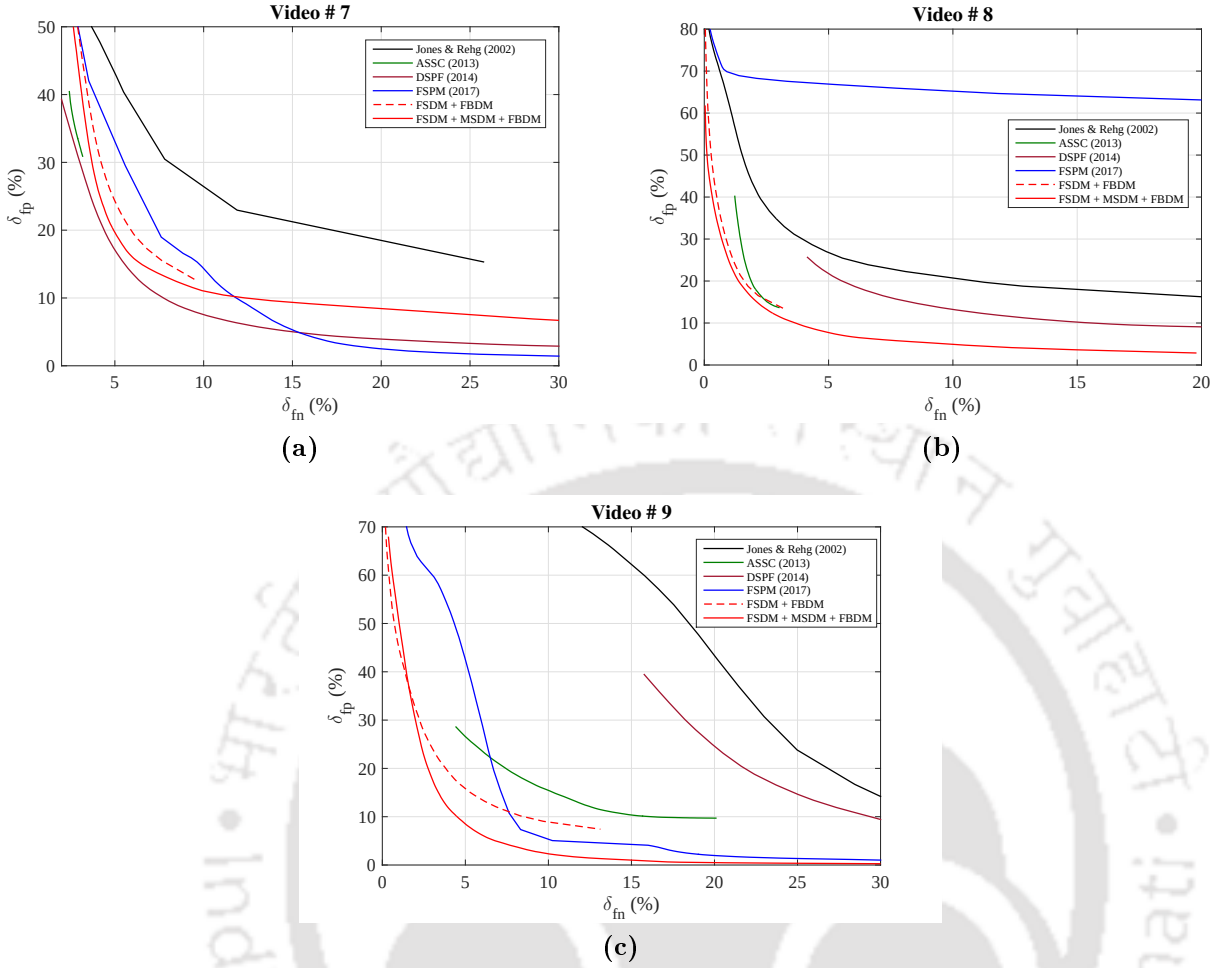


Figure 4.13: ROC plots for different videos (7–9).

produces more false negatives as compared to our proposed method. In order to compare the SDDMA method with the proposed method, it is trained with N_{TF} number of labelled initial frames of 9 videos (In total, $9 \times N_{TF}$ frames). However, SDDMA fails significantly in detecting true skin regions. It produces highest false negatives among all the benchmark methods. The DSPF method can give comparatively better results than the standard SPM, and it gives a discriminative space-based skin map. The DSPF method mostly relies on SPM derived from global training samples, and thus it is not adaptive to local environmental conditions of an image.

As human facial colour almost resembles the overall skin colour of a person, face detection-based skin model adaptation technique can give superior skin detection performance. The ASSC method uses adaptive seeds for skin region growing. The adaptive seeds are derived from a

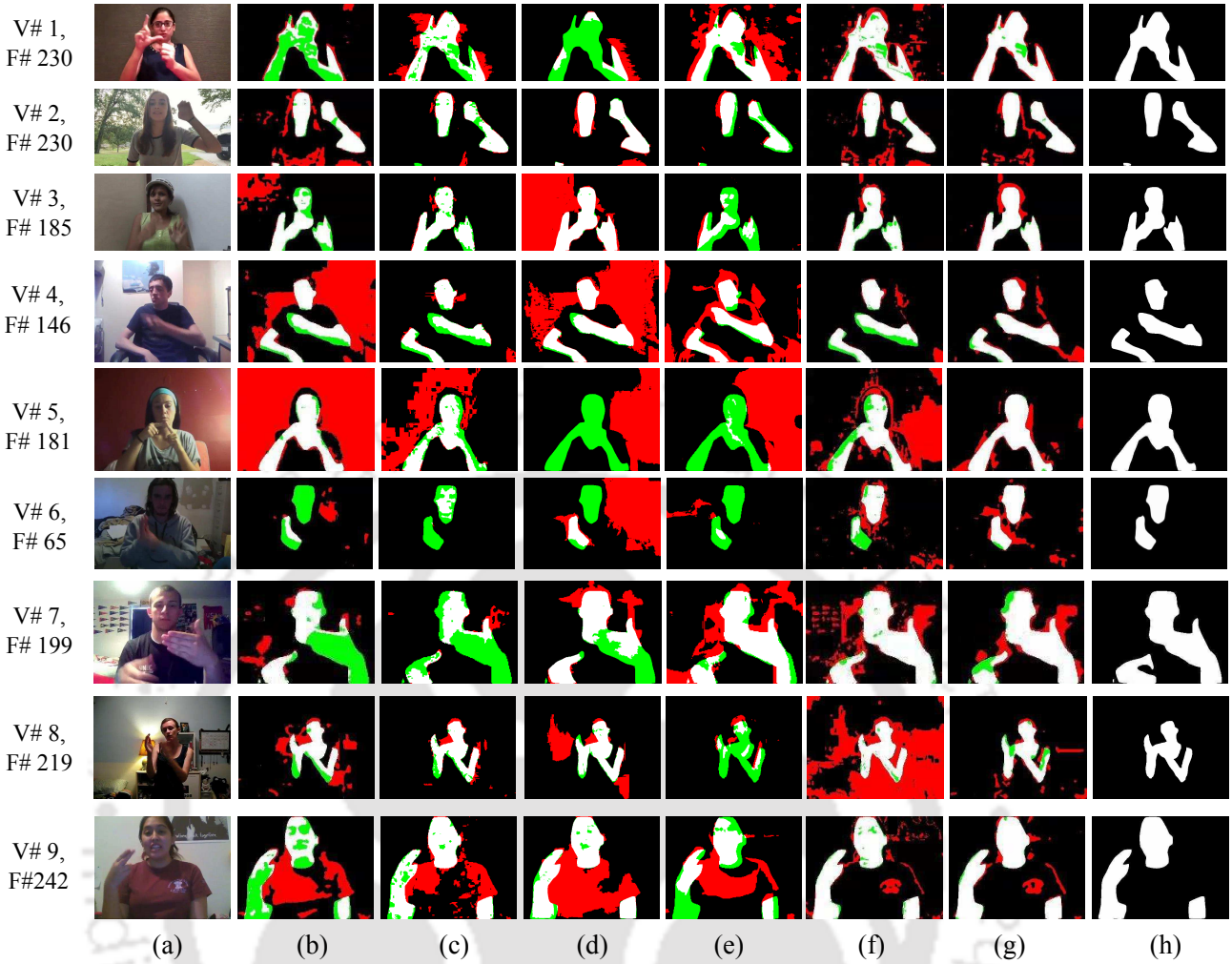


Figure 4.14: Comparative detection results for minimum $\delta_{t,avg}$: (a) Original frames, (b) Jones and Rehg [13], (c) ASSC [21], (d) DSPF [9], (e) SASC [22], (f) FSPM [23], (g) Proposed method , and (h) Groundtruth. Here, white colour represents true positive, black colour represents true negative, red colour represents false positive, and green colour represents false negative.

local skin model (derived from facial skin pixels). The ASSC method relies on a standard SPM for region growing, and hence, it is unable to detect many true skin pixels. So, this method produces higher false negative errors than the proposed method. The FSPM-based method gives better detection accuracy by utilizing image pixel distribution information to derive a local SPM (LSPM). The LSPM is later fused with the original or global SPM to get the FSPM. The FSPM is able to detect more true skin pixels as compared to other benchmark methods. In our method, the combination of FSDM and FBDM gives a static skin detection model for videos. The incorporation of MSDM makes the skin detection model adaptive to local illumination changes, and a dynamic skin detection model for videos is obtained. The static

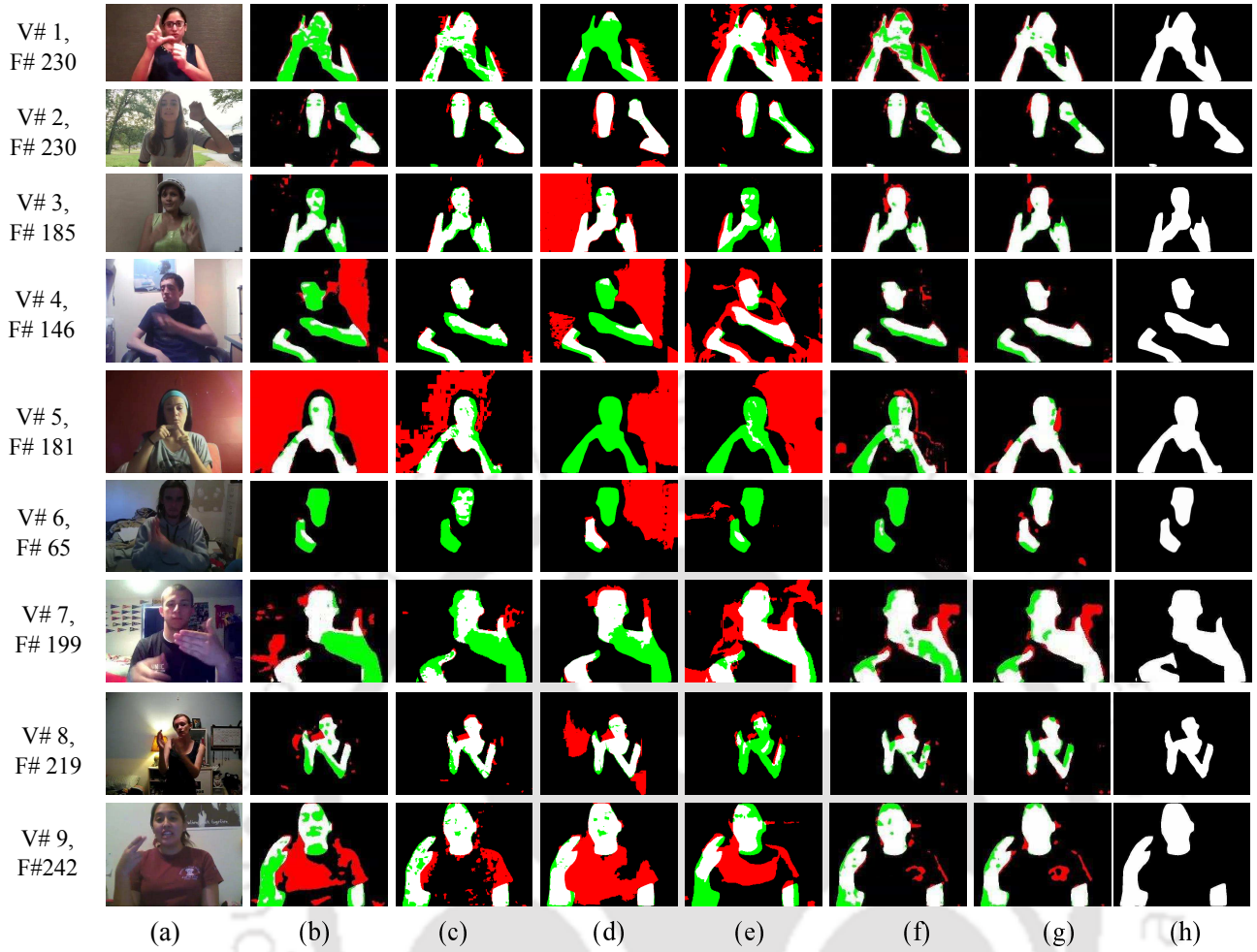


Figure 4.15: Comparative detection results for maximum accuracy: (a) Original frames, (b) Jones and Rehg [13], (c) ASSC [21], (d) DSPF [9], (e) SASC [22], (f) FSPM *et al.* [23], (g) Proposed method , and (h) Groundtruth. Here, white colour represents true positive, black colour represents true negative, red colour represents false positive, and green colour represents false negative.

model is prone to produce more false rejections as compared to its dynamic counterpart. The dynamic adaptation of the proposed MSDM for local illumination changes makes the proposed system more robust to unconstrained illumination and background conditions. In Fig. 4.12–4.13, comparative ROC plots are shown for different videos. In Fig. 4.14–4.15, the detection results are shown for all the test videos. It is evident from all the experiential results that the proposed method can detect skin regions in a video more accurately than the existing standard methods even in the presence of local colour deformations.

4.4 Summary

Segmentation of skin regions is a necessary step for various vision-based Human-Computer Interaction (HCI) applications. However, detection of skin regions in videos is a challenging problem when the shading of other body parts causes local chromatic variations even if the global illumination remains constant. In this paper, a skin detection algorithm is proposed which dynamically adapts to local skin colour variations. The proposed skin detection method has three major components – a facial skin pixel distribution model (FSDM) for user-specific skin modelling, a video specific local background distribution model (LBDM), and a moving skin-pixel distribution model (MSDM) for detection of skin regions affected by local colour deformations due to the motion of body parts. The FSDM is derived by using a set of facial pixels from initial frames of a video. The LBDM is derived from the FSDM and a global background distribution model (GBDM). A modified version of the Bhattacharyya distance (MBD) is employed to measure similarity between two distribution models. Subsequently, a fusion-based background distribution model (FBDM) is derived by utilising GBDM and LBDM. The MSDM gives the distribution of moving skin pixels. The final skin detection model is derived by utilising FSDM, MSDM and FBDM. As the MSDM is updated at every keyframe, the proposed skin model becomes adaptive to local illumination changes. From the experimental results, it is evident that the use of MBD produces comparatively less detection errors. One important research direction would be the detection of skin regions affected by both local and global illumination variations.

5

Conclusions and Future Work

There is a humbling amount of past works on skin segmentation. This dissertation describes our proposed skin segmentation algorithms for various uncontrolled environments, such as presence of skin-like colours in background, non-uniform and dynamic illumination. In our method, multiple local skin model adaptation schemes have been proposed by utilizing chromatic and textural features. Also, a dynamic region growing scheme has been proposed for additional reduction in detection errors. This chapter reflects on these contributions, discusses future work for skin detection, and concludes. We hope that the future skin detection algorithm developers find our contributions useful, and get benefited from our local skin model adaption schemes without having to reinvent their own.

5.1 Summary

In the beginning of this thesis, challenges faced by computer vision community in detecting skin regions are mentioned. These challenges include presence of skin-like colours in background, non-uniform illumination (static), and varying illumination (dynamic). In this research work, we have proposed multiple frameworks for skin detection using chromatic and textural features to overcome some of the above-mentioned challenges. The main goal of this dissertation is to obtain skin detection methods which can be used as a pre-processing step for different HCI applications in an uncontrolled environment.

In summary, this thesis addresses the following issues:

- Extraction of image specific discriminative features between skin and non-skin regions of an image for skin detection in presence of skin-like background colours.
- Development of an image specific discriminant analysis approach to handle marginal inter-class overlapping of skin pixel samples of the training dataset.
- Development of a skin region growing approach for an image by using a local adaptation scheme.
- Development of a local skin model adaptation scheme for a global skin detection model by utilizing image pixel distribution information.
- Development of a video-specific background model for skin detection in videos.
- Development of a detection scheme for skin regions having local colour deformation due to motion.

The main contributions of the thesis can be summarized as follows:

- An image specific discriminative feature extraction method is proposed for better discrimination between skin and non-skin regions.
- A measure termed as *skin-chroma entropy* is proposed to estimate false acceptance error.
- A dynamic region growing (DRG) method is proposed for additional reduction in detection errors. The extent of region growing depends on skin-chroma entropy value of an image.

- An image pixel distribution information is employed for deriving an adaptive skin model.
- A video specific background pixel distribution model is proposed.
- A moving pixel distribution model (MSDM) is proposed for moving skin regions.
- A modified version of the Bhattcharyya distance is proposed for better similarity measurement between an unknown skin distribution and a reference skin distribution.
- A skin detection model is proposed by utilizing the Facial Skin Distribution Model (FSDM), the Moving Skin pixel Distribution Model (MSDM), and the Fusion-based Background Distribution Model (FBDM) to detect skin regions in videos.

The summary of all the chapters of this dissertation is highlighted as follows:

i) **Chapter 1** provides a brief description of a typical skin detection method and its different components, such as usable colour spaces and classifiers. It also reviews existing skin detection methods. The existing skin detection techniques can be grouped into following two main categories: a) static framework-based approaches, and b) dynamic framework-based approaches. Static framework-based approaches were primarily developed for both images and videos having static background and fixed illumination conditions. On the other-hand, dynamic framework-based methods were primarily developed for videos when illumination varies among the frames either globally or locally. The static framework-based methods can be further grouped into following three sub-categories based on their classification approaches: a) explicit boundary specification of colour components, b) skin classification using parametric modelling of skin colour distribution, and c) skin classification using non-parametric modelling of skin colour distribution. Explicit boundary specification of colour components is the simplest and the fastest approaches among all these approaches. However, the method cannot produce satisfactory results due to lack of generalization and rigidity in structure. Parametric modelling-based methods use less data for generalization, it and they are computationally less complex than non-parametric modelling-based approaches. Non-parametric approaches are either based on a histogram-based pixel distribution approximation or different supervised classifiers, such as non-liner SVM, ANN, Random Forest. Supervised classifiers need labelled training data, and hence, any inter-class similarity in training data can result in an inaccurate modelling of input

data. Also, the training procedures are computationally expensive for these classifiers than the parametric modelling. All these approaches are elaborately discussed in **Chapter 1**.

In **Chapter 1**, challenges of skin detection approaches are also discussed. Major challenges include - presence of skin like colours in background, chromatic deformation in skin appearance due to non-uniform or time-varying illuminations. Finally, **Chapter 1** is ended up with motivation, objectives, and the organization of the thesis.

ii) **Chapter 2** focuses on developing a method for extracting discriminative features between skin and non-skin regions of an image. This is accomplished by extracting two sets of sample pixels, out of which one contains mostly skin pixels and another contains mostly non-skin pixels. Subsequently, an adaptive discriminative analysis (ADA) has been proposed to find the discriminative features using these two sample sets. A skin region growing scheme with dynamic adaptation to image properties is proposed for additional reduction in detection errors. The performance of the proposed method is evaluated using standard databases, and it is observed that the proposed method can provide better result than state-of-the-art methods.

iii) **Chapter 3** explores an use of image pixel distribution information in the context of skin detection. A GMM is used to approximate the pixel distribution of an image, which is also termed as image distribution model (IDM) in a given colour space. Subsequently, a local skin detection model is derived by using the similarity between the IDM and a reference skin distribution model. The reference skin model is derived by using a set of skin pixels belonging to the facial region of a person. Finally, a fusion-based skin probability map (FSPM) for the image is obtained by fusing an SPM derived from the local skin detection model with another SPM derived from a global skin detection model. The proposed dynamic region growing (DRG) algorithm is applied on the FSPM for a more accurate skin detection. The efficacy of the proposed method is tested by using a set of standard datasets. Experimental results show that the proposed method can detect skin regions in images more accurately than the state-of-the-art methods.

iv) **Chapter 4** describes a skin detection method for videos under local colour deformation of skin regions. Local skin colour deformation may occur due to non-uniform illumination caused by the motion of hands or other body parts. The proposed method updates the skin

model dynamically at the frames having significant chromatic changes. The performance of this method is evaluated on a set of manually annotated sign language videos. Experimental results show that the proposed method can significantly improve the overall skin detection accuracy in videos under uncontrolled background and lighting conditions.

5.2 Future Work

Our proposed work addressed a number of existing issues of skin colour detection and segmentation. It also points to certain areas which could benefit from further research.

- First of all, my work did not explicitly address the issues of detecting skin regions in a video when the background is dynamic.
- Accurate segmentation of skin regions of video frames under a combined effect of global and local illumination changes is a major challenge in real-life applications, and this could be an important research direction.
- Our proposed methods used contextual information of an image or a video to obtain self-adaptive skin models. However, these models depend on manually selected chromatic and textural features. On the other hand, deep learning-based architectures are capable of extracting low level to high level features automatically for difference computer vision applications. Therefore, future work could be motivated towards an incorporation of the proposed local adaptation schemes onto a deep-learning-based architecture.

List of Publications

Journal Publications

1. **B. K. Chakraborty**, M.K. Bhuyan, and S. Kumar, "Combining image and global pixel distribution model for skin colour segmentation", *Pattern Recognition Letters*, vol. 88, pp. 33—40, 2017

2. **B. K. Chakraborty**, Debajit Sarma, M.K. Bhuyan, Karl F. MacDorman, "Review of Constraints on Vision-based Gesture Recognition for Human-Computer Interaction", *IET Computer Vision*, vol. 12, no. 1, pp. 3–15, 2018.

Manuscripts under Review

1. **B. K. Chakraborty**, M.K. Bhuyan, S. Kumar, "Image Specific Discriminative Feature Extraction for Skin Segmentation", major revision in *Journal of Visual Communication and Image Representation*, Elsevier (Revised paper submitted).

2. **B. K. Chakraborty**, M.K. Bhuyan, Karl F. MacDorman, "Skin Segmentation in Videos under Locally Varying Illumination", communicated to *IEEE Trans. Circuits and Systems for Video Technology* .

Conference Publications

1. **B. K. Chakraborty**, M.K. Bhuyan, S. Kumar, "Fusion-based skin detection using image distribution model", *in Proc. Tenth Indian Conf. Computer Vision , Graphics and Image Processing (ICVGIP)*, Guwahati, Dec. 2016, pp. 67:1—67:8.

2. **B. K. Chakraborty**, M.K. Bhuyan and S. Kumar, "Adaptive propagation-based skin segmentation method for color images", *in Proc. 2016 Twenty Second National Conf. Communication (NCC)*, Guwahati, 2016, pp. 1–6.

3. **B. K. Chakraborty**, M.K. Bhuyan and S. Kumar, "A Weighted Skin Probability Map for skin color segmentation", *in Proc. IEEE Int. Conf. Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, 2016, pp. 2133–2136.

4. **B. K. Chakraborty** and M.K. Bhuyan, "Skin segmentation using Possibilistic Fuzzy C-means clustering in presence of skin-colored background", *in Proc. IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, Trivandrum, 2015, pp. 246–250.

Bibliography

[1] A. Elgammal, C. Muang, and D. Hu, *Skin Detection - a Short Tutorial.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 1218–1224.

[2] S. Tsekeridou and I. Pitas, "Facial feature extraction in frontal views using biometric analogies," in *Proc. 9th European Signal Process. Conf (EUSIPCO 1998)*, Sept 1998, pp. 1–4.

[3] F. Solina, P. Peer, B. Batagelj, and S. Juvan, "15 seconds of fame - an interactive, computer-vision based art installation," in *Proc. 7th Int. Conf. Control, Auto., Robot. and Vis.*, vol. 1, Dec 2002, pp. 198–204.

[4] R.-L. Hsu, M. Abdel-Mottaleb, and A. Jain, "Face detection in color images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 696–706, 2002.

[5] G. Kukharev and A. Nowosielski, "Fast and efficient algorithm for face detection in colour images," *MG and V*, vol. 13, no. 4, pp. 377–399, 2004.

[6] A. Cheddad, J. Condell, K. Curran, and P. M. Kevitt, "A skin tone detection algorithm for an adaptive approach to steganography," *Signal Processing*, vol. 89, no. 12, pp. 2465–2478, 2009.

[7] Y.-H. Chen, K.-T. Hu, and S.-J. Ruan, "Statistical skin color detection method without color transformation for real-time surveillance systems," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1331–1337, 2012.

[8] M. Kawulok, J. Nalepa, and J. Kawulok, *Skin Detection and Segmentation in Color Images.* Dordrecht: Springer Netherlands, 2014, pp. 329–366.

[9] M. Kawulok, J. Kawulok, and J. Nalepa, "Spatial-based skin detection using discriminative skin-presence features," *Pattern Recognit. Lett.*, vol. 41, no. 0, pp. 3–13, 2014.

[10] W. R. Tan, C. S. Chan, P. Yogarajah, and J. Condell, "A fusion approach for efficient human skin detection," *IEEE Trans. Ind. Inform.*, vol. 8, no. 1, pp. 138–147, 2012.

[11] M.-J. Seow, D. Valaparla, and V. K. Asari, "Neural network based skin color model for face detection," in *Proc. 32nd App. Imagery Pattern Recognit. Workshop*, Oct 2003, pp. 141–145.

[12] G. Yang, H. Li, L. Zhang, and Y. Cao, "Research on a skin color detection algorithm based on self-adaptive skin color model," in *Proc. Int. Conf. Communications and Intelligence Information Security (ICCIIS)*, 2010, pp. 266–270.

[13] M. J. Jones and J. Rehg, "Statistical color models with application to skin detection," *Int. J. Comput. Vis.*, vol. 46, no. 1, pp. 81–96, 2002.

[14] M. Kawulok, "Fast propagation-based skin regions segmentation in color images," in *Proc. 10th IEEE Int. Conf. and Workshops Autom. Face and Gesture Recognit. (FG)*, April 2013, pp. 1–7.

[15] M. Kawulok, J. Kawulok, J. Nalepa, and B. Smolka, "Self-adaptive algorithm for segmenting skin regions," *EURASIP J. Adv. Signal Process.*, vol. 2014, pp. 1–22, 2014.

[16] R. Hettiarachchi and J. Peters, "Multi-manifold-based skin classifier on feature space voronoï regions for skin segmentation," *J. Vis. Commun. Image R.*, vol. 41, pp. 123–139, 2016.

[17] R. Khan, A. Hanbury, and J. Stoettinger, "Skin detection: a random forest approach," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sept 2010, pp. 4613–4616.

[18] D. J. Sawicki and W. Miziolek, "Human colour skin detection in CMYK colour space," *IET Image Process.*, vol. 9, no. 9, pp. 751–757, 2015.

[19] M. Kawulok, *Dynamic skin detection in color images for sign language recognition.* Berlin, Heidelberg: Springer Berlin Heidelberg, July 2008, ch. Proc. Int. Conf. Image and Signal Process. (ICISP), pp. 112–119.

[20] P. Yogarajah, J. Condell, K. Curran, A. Cheddad, and P. McKevitt, "A dynamic threshold approach for skin segmentation in color images," in *Proc. 17th IEEE Int. Conf. Image Process.*, Sept 2010, pp. 2225–2228.

[21] M. Kawulok, J. Kawulok, J. Nalepa, and M. Papiez, "Skin detection using spatial analysis with adaptive seed," in *Proc. IEEE Int. Conf. Image Process.*, Sept 2013, pp. 3720–3724.

[22] Y. Lei, W. Yuan, H. Wang, Y. Wenhu, and W. Bo, "A skin segmentation algorithm based on stacked autoencoders," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 740–749, 2017.

[23] B. K. Chakraborty, M. Bhuyan, and S. Kumar, "Combining image and global pixel distribution model for skin colour segmentation," *Pattern Recognit. Lett.*, vol. 88, pp. 33–40, 2017.

[24] B. Fink, K. Grammer, and P. J. Matts, "Visible skin color distribution plays a role in the perception of age, attractiveness, and health in female faces," *Evolution and Human Behavior*, vol. 27, no. 6, pp. 433–442, 2006.

[25] J. Han, G. Award, A. Sutherland, and H. Wu, "Automatic skin segmentation for gesture recognition combining region and support vector machine active learning," in *Proc. 7th Int. Conf. Autom. Face and Gesture Recognit.*, Apr. 2006, pp. 237–242.

[26] S. Bilal, R. Akmeliawati, M. J. E. Salami, and A. A. Shafie, "Dynamic approach for real-time skin detection," *J. Real-Time Image Process.*, vol. 10, no. 2, pp. 371–385, 2015.

[27] J.-S. Lee, Y.-M. Kuo, P.-C. Chung, and E.-L. Chen, "Naked image detection based on adaptive and extensible skin color model," *Pattern Recognit.*, vol. 40, no. 8, pp. 2261–2270, 2007.

[28] Q. Zhu, C.-T. Wu, K.-T. Cheng, and Y.-L. Wu, "An adaptive skin model and its application to objectionable image filtering," in *Proc. 12th Annual ACM Int. Conf. Multimedia*, ser. MULTIMEDIA '04, 2004, pp. 56–63.

[29] H. Kruppa, M. A. Bauer, and B. Schiele, *Skin patch detection in real-world images.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 109–116.

[30] M. Abdullah-Al-Wadud and O. Chae, "Region-of-interest selection for skin detection based applications," in *Proc. Int. Conf. Convergence Information Technology (ICCIT)*, Nov 2007, pp. 1999–2004.

[31] M.-J. Chen, M.-C. Chi, C.-T. Hsu, and J.-W. Chen, "RoI video coding based on H.263+ with robust skin-color detection technique," in *Proc. IEEE Int. Conf. Consum. Electron.*, June 2003, pp. 44–45.

[32] B. Choi, B. Chung, and J. Ryou, "Adult image detection using Bayesian decision rule weighted by svm probability," in *Proc. Fourth Int. Conf. Computer Sciences and Convergence Information Technology*, Nov 2009, pp. 659–662.

[33] D. Forsyth and M. Fleck, "Automatic detection of human nudes," *Int. J. Computer Vision*, vol. 32, no. 1, pp. 63–77, 1999.

[34] M. M. Fleck, D. A. Forsyth, and C. Bregler, *Finding naked people.* Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 593–602.

[35] Z. H. Al-Tairi, R. W. Rahmat, M. I. Saripan, and P. S. Sulaiman, "Skin segmentation using YUV and RGB color spaces," *J. Info. Process. Sys.*, vol. 10, pp. 283–299, 2014.

[36] K. Saifullah, *Applied computational intelligence and soft computing in engineering.* Hershey, PA, USA: IGI Global, 2018, [Online; available 22 Sept. 2017].

[37] D. Chai and A. Bouzerdoum, "A Bayesian approach to skin color classification in YCbCr color space," in *Proc. TENCON 2000*, vol. 2, 2000, pp. 421–424.

[38] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 2002, pp. 289–292.

[39] T. Mahmoud, "A new fast skin color detection technique," *World Acad. Sci. Eng. Technol.*, vol. 43, pp. 501–505, 2008.

[40] J. Basilio, G. Torres, G. Perez, L. Medina, and H. Meana, "Explicit image detection using YCbCr space color model as skin detection," in *Proc. American Conf. on Applied Mathematics and the Fifth WSEAS Int. Conf. on Computer Engineering and Applications, World Scientific and Engineering Academy and Society (WSEAS)*, 2011, pp. 123–128.

[41] J. Kovac, P. Peer, and F. Solina, "Human skin color clustering for face detection," in *Proc. EUROCON 2003, Computer as a Tool*, vol. 2, 2003, pp. 144–148.

[42] B. C. Ennehar, O. Brahim, and T. Hicham, "Skin segmentation using YUV and RGB color spaces," *INFOCOMP J. Comput. Sc.*, vol. 9, pp. 1–10, 2010.

[43] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," *Signal Process.: Image Commun.*, vol. 12, no. 3, pp. 263–281, 1998.

[44] A. Albiol, L. Torres, and E. J. Delp, "Optimum color spaces for skin detection," in *Proc. IEEE Int. Conf. Image Process.*, vol. 1, 2001, pp. 122–124.

[45] B. Muhammad and S. A. R. Abu-Bakar, "A hybrid skin color detection using hsv and ycgcr color space for face detection," in *Proc. IEEE Int. Conf. Signal and Image Process. App. (ICSIPA)*, Oct 2015, pp. 95–98.

[46] B. Zarit, B. Super, and F. Quek, "Comparison of five color models in skin pixel classification," in *Proc. ICCV'99 Int. Workshop on Recognit., Anal., and Tracking of Faces and Gestures in Real-Time Syst.*, 1999, pp. 58–63.

[47] R. P. Schumeyer and K. E. Barner, "Color-based classifier for region identification in video," in *Proc. SPIE 3309, Visual Comm. and Image Process.*, Jan 1999, pp. 189–200.

[48] D. A. Brown, I. Craw, and J. Lewthwaite, "A som based approach to skin detection with application in real time systems," in *Proc. British Mach. Vis. Conf.* BMVA Press, 2001, pp. 51.1–51.10.

[49] J. M. Chaves-González, M. A. Vega-Rodríguez, J. A. Gómez-Pulido, and J. M. Sánchez-Pérez, "Detecting skin in face recognition systems: A colour spaces study," *Digital Signal Process.*, vol. 20, no. 3, pp. 806–823, 2010.

[50] J. Brand and J. Mason, "A comparative assessment of three approaches to pixel-level human skin-detection," in *Proc. 15th Int. Conf. Pattern Recognit.*, vol. 1, 2000, pp. 1056–1059.

[51] G. Gomez, M. Sanchez, and L. Enrique Sucar, *On selecting an appropriate colour space for skin detection.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 69–78.

[52] J. Yang and A. Waibel, "A real-time face tracker," in *Proc. 3rd IEEE Work. App. Comput. Vis. (WACV)*, ser. WACV '96, 1996, pp. 142–147.

[53] J. Yang, W. Lu, and A. Waibel, "Skin-color modeling and adaptation," in *Computer Vision — ACCV'98*, ser. Lecture Notes in Computer Science, R. Chin and T.-C. Pong, Eds. Springer Berlin Heidelberg, Jul. 1997, vol. 1352, pp. 687–694.

[54] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its applications in image and video databases," in *Proc. SPIE 3656, Storage and Retrieval for Image and Video Databases VII*, vol. 3656, Dec. 1998, pp. 458–466.

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

[56] C. Chen and S. P. Chiang, "Detection of human faces in colour images," *IEE Proc. Vision, Image and Signal Process.*, vol. 144, no. 6, pp. 384–388, 1997.

[57] J. Karlekar and U. B. Desai, "Finding faces in color images using wavelet transform," in *Proc. 10th Int. Conf. Image Anal. and Process.*, 1999, pp. 1085–1088.

[58] L. Chen, J. Zhou, Z. Liu, W. Chen, and G. Xiong, "A skin detector based on neural network," in *Proc. IEEE Int. Conf. Commun., Circuits and Syst. and West Sino Expositions*, vol. 1, June 2002, pp. 615–619.

[59] A. Zaidan, N. Ahmad, H. A. Karim, M. Larbani, B. Zaidan, and A. Sali, "Image skin segmentation based on multi-agent learning Bayesian and neural network," *Eng. Appl. Artif. Intell.*, vol. 32, no. Supplement C, pp. 136–150, 2014.

[60] E. Casiraghi, R. Lanzarotti, and G. Lipori, *A face detection system based on color and support vector machines*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 113–120.

[61] R. Schettini, C. Brambilla, C. Cusano, and G. Ciocca, "On the detection of pornographic digital images," in *Proc. SPIE*, vol. 5150, 2003, pp. 2150–2113.

[62] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[63] T. K. Ho, "Random decision forests," in *Proc. 3rd Int. Conf. Doc. Anal. and Recognit.*, ser. ICDAR '95. Washington, DC, USA: IEEE Computer Society, 1995, pp. 278–282.

[64] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[65] S. Phung, A. Bouzerdoum, and S. Chai, D., "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 1, pp. 148–154, 2005.

[66] S. J. Schmugge, S. Jayaram, M. C. Shin, and L. V. Tsap, "Objective evaluation of approaches of skin detection using roc analysis," *Computer Vision and Image Understanding*, vol. 108, no. 1, pp. 41–51, 2007.

[67] J. Nalepa, T. Grzejszczak, and M. Kawulok, "Wrist localization in color images for hand gesture recognition," in *Man-Machine Interactions 3*, ser. Adv. Intell. Syst. and Comput., D. A. Gruca, T. Czachórski, and S. Kozielski, Eds. Springer International Publishing, 2014, vol. 242, pp. 79–86.

[68] T. K. Kim, S. F. Wong, and R. Cipolla, "Tensor canonical correlation analysis for action classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, June 2007, pp. 1–8.

[69] T. K. Kim and R. Cipolla, "Canonical correlation analysis of video volume tensors for action categorization and detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 8, pp. 1415–1428, 2009.

[70] P. K. Pisharady, P. Vadakkepat, and A. P. Loh, "Attention based detection and recognition of hand postures against complex backgrounds," *Int. J. Comput. Vis.*, vol. 101, no. 3, pp. 403–419, 2013.

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

[72] A. Nikolaidis and I. Pitas, "Robust watermarking of facial images based on salient geometric pattern matching," *IEEE Trans. Multimedia*, vol. 2, no. 3, pp. 172–184, Sep 2000.

[73] K. B. Shaik, P. Ganesan, V. Kalist, B. Sathish, and J. M. M. Jenitha, "Comparative study of skin color detection and segmentation in HSV and YCbCr color space," *Procedia Computer Science*, vol. 57, no. Supplement C, pp. 41–48, 2015.

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

[75] X. Wang, X. Zhang, and J. Yao, "Skin color detection under complex background," in *Proc. Int Conf. Mechatronic Science, Electric Engineering and Computer (MEC)*, Aug 2011, pp. 1985–1988.

[76] M.-H. Yang and N. Ahuja, "Gaussian mixture model for human skin color and its applications in image and video databases vii," vol. 3656, 1999, pp. 458–466.

[77] H. Greenspan, J. Goldberger, and I. Eshet, "Mixture model for face-color modeling and segmentation," *Pattern Recognit. Lett.*, vol. 22, no. 14, pp. 1525–1536, 2001.

[78] T. S. Caetano, S. D. Olabarriaga, and D. A. Barone, "Do mixture models in chromaticity space improve skin detection?" *Pattern Recognit.*, vol. 36, no. 12, pp. 3019–3021, 2003.

[79] Q. Huynh-Thu, M. Meguro, and M. Kaneko, "Skin-color extraction in images with complex background and varying illumination," in *Proc. Sixth IEEE Work. App. Comput. Vis.*, 2002, pp. 280–285.

[80] A. A. Abin, M. Fotouhi, and S. Kasaei, "A new dynamic cellular learning automata-based skin detector," *Multimedia Systems*, vol. 15, no. 5, pp. 309–323, 2009.

[81] A. Conci, E. Nunes, J. J. Pantrigo, and A. Sanchez, "Comparing color and texture-based algorithms for human skin detection," in *Proc. ICEIS*, Aug 2008, pp. 1985–1988.

[82] M. Kawulok, "Texture analysis for skin probability maps refinement," in *Pattern Recognit.*, ser. Lecture Notes in Computer Science, J. Carrasco-Ochoa, J. Martínez-Trinidad, J. Olvera López, and K. Boyer, Eds. Springer Berlin Heidelberg, 2012, vol. 7329, pp. 75–84.

[83] Z. Jiang, M. Yao, and W. Jiang, "Skin detection using color, texture and space information," in *Proc. Fourth Int. Conf. Fuzzy Syst. and Knowledge Discovery*, vol. 3, Aug 2007, pp. 366–370.

[84] H.-M. Sun, "Skin detection for single images using dynamic skin color modeling," *Pattern Recognit.*, vol. 43, no. 4, pp. 1413–1420, 2010.

[85] P. Ng and C.-M. Pun, "Skin color segmentation by texture feature extraction and k-mean clustering," in *Proc. 3rd Int. Conf. Comput. Intell., Commun. Syst. and Netw. (CICSyN)*, July 2011, pp. 213–218.

[86] J. Ruiz-del Solar and R. Verschae, "Skin detection using neighborhood information," in *Proc. 6th IEEE Int. Conf. Auto. Face and Gesture Recognit.*, May 2004, pp. 463–468.

[87] M. Kawulok, "Energy-based blob analysis for improving precision of skin segmentation," *Multimedia Tools and Applications*, vol. 49, no. 3, pp. 463–481, 2010.

[88] L. Ikonen and P. Toivanen, "Distance and nearest neighbor transforms on gray-level surfaces," *Pattern Recognit. Lett.*, vol. 28, no. 5, pp. 604–612, 2007.

[89] J. Fritsch, S. Lang, A. Kleinehagenbrock, G. A. Fink, and G. Sagerer, "Improving adaptive skin color segmentation by incorporating results from face detection," in *Proc. IEEE Int. Workshop Robot and Human Interactive Comm.*, 2002, pp. 337–343.

[90] D. Chai, S. L. Phung, and A. Bouzerdoum, "A Bayesian skin/non-skin color classifier using non-parametric density estimation," in *Proc. ISCAS*, vol. 2, May 2003, pp. II–464–II–467 vol.2.

[91] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 862–877, 2004.

[92] S. Shirali-Shahreza and M. E. Mousavi, "A new Bayesian classifier for skin detection," in *Proc. 3rd Int. Conf. Innovative Comput. Info. and Control*, June 2008, pp. 172–172.

[93] S. Khan, D. Bailey, G. S. Gupta, and S. Demidenko, "Adaptive classifier for robust detection of signing articulators based on skin colour," in *Proc. Sixth IEEE Int. Symp. Elect. Design, Test and App.*, Jan 2011, pp. 259–262.

[94] D. R. Musicant and A. Feinberg, "Active set support vector regression," *IEEE Trans. Neural Netw.*, vol. 15, no. 2, pp. 268–275, March 2004.

[95] L. Duan, Z. Lin, J. Miao, and Y. Qiao, *A method of human skin region detection based on PCNN.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 486–493.

[96] S. Bianco, F. Gasparini, and R. Schettini, "Adaptive skin classification using face and body detection," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4756–4765, Dec 2015.

[97] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Adaptive skin color modeling using the skin locus for selecting training pixels," *Pattern Recognit.*, vol. 36, no. 3, pp. 681–690, 2003.

[98] L. Sigal, S. Sclaroff, and V. Athitsos, "Skin color-based video segmentation under time-varying illumination," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 7, pp. 862–877, 2004.

[99] N. Habili, C.-C. Lim, and A. Moini, "Segmentation of the face and hands in sign language video sequences using color and motion cues," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 8, pp. 1086–1097, 2004.

[100] G. Awad, J. Han, and A. Sutherland, "A unified system for segmentation and tracking of face and hands in sign language recognition," in *Proc. Int. Conf. Pattern Recognit.*, vol. 1, 2006, pp. 239–242.

[101] J. Han, G. Awad, and A. Sutherland, "Automatic skin segmentation and tracking in sign language recognition," *IET Comput. Vis.*, vol. 3, no. 1, pp. 24–35, 2009.

[102] L. Liu, N. Sang, S. Yang, and R. Huang, "Real-time skin color detection under rapidly changing illumination conditions," *IEEE Trans. Consum. Electron.*, vol. 57, no. 3, pp. 1295–1302, 2011.

[103] B. K. Chakraborty and M. K. Bhuyan, "Skin segmentation using possibilistic fuzzy c-means clustering in presence of skin-colored background," in *Proc. IEEE Recent Adv. Intel. Comput. Systs (RAICS)*, Dec 2015, pp. 246–250.

[104] D. Xu, Y.-L. Chen, X. Wu, Y. Ou, and Y. Xu, "Integrated approach of skin-color detection and depth information for hand and face localization," in *Proc. IEEE Int. Conf. Robotics and Biomimetics (ROBIO)*, Dec 2011, pp. 952–956.

[105] P. Trindade, J. Lobo, and J. Barreto, "Hand gesture recognition using color and depth images enhanced with hand angular pose data," in *Proc. IEEE Conf. Multisensor Fusion and Integration for Intell. Syst. (MFI)*, Sept 2012, pp. 71–76.

[106] C. Dumitrescu and I. Dumitrache, "Human skin detection using texture information and vector processing techniques by neural networks," in *Adv. in Intell. Control Syst. and Comput. Sc.*, ser. Adv.. Intell. Syst. and Comput., L. Dumitrache, Ed. Springer Berlin Heidelberg, 2013, vol. 187, pp. 59–75.

[107] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, June 2015, pp. 3431–3440.

[108] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.

[109] L. Ladicky, C. Russell, P. Kohli, and P. H. Torr, "Associative hierarchical CRFs for object class image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, October 2009.

[110] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3d brain image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1744–1757, Oct 2010.

[111] R. Mottaghi, X. Chen, X. Liu, N. G. Cho, S. W. Lee, S. Fidler, R. Urtasun, and A. Yuille, "The role of context for object detection and semantic segmentation in the wild," in *Proc. IEEE Cong. Comput. Vis. Pattern Recognit.*, June 2014, pp. 891–898.

[112] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, ser. ICCV '15. Washington, DC, USA: IEEE Computer Society, 2015, pp. 1529–1537.

[113] B. Li, X. Xue, and J. Fan, "A robust incremental learning framework for accurate skin region segmentation in color images," *Pattern Recognit.*, vol. 40, no. 12, pp. 3621–3632, 2007.

[114] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 6, pp. 641–647, 1994.

[115] N. Pal, K. Pal, J. Keller, and J. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 4, pp. 517–530, 2005.

[116] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man and Cybern.*, vol. 9, no. 1, pp. 62–66, 1979.

[117] B. K. Chakraborty, M. K. Bhuyan, and S. Kumar, "Fusion-based skin detection using image distribution model," in *Proc. Tenth Indian Conf. Comput. Vis. Graphics and Image Process.*, ser. ICVGIP '16, 2016, pp. 67:1–67:8.

[118] ——, "Adaptive propagation-based skin segmentation method for color images," in *Proc. National Conf. Comm.*, Sept 2016, pp. 94–99.

[119] T. Xu, Y. Wang, and Z. Zhang, "Pixel-wise skin colour detection based on flexible neural tree," *IET Image Process.*, vol. 7, no. 8, pp. 751–761, 2013.

[120] C.-F. Juang, S.-H. Chiu, and S.-J. Shiu, "Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation," *IEEE Trans. Syst., Man, Cybern., Part A: Syst. and Humans*, vol. 37, no. 6, pp. 1077–1087, 2007.

[121] C.-F. Juang, C.-M. Chang, J.-R. Wu, and D. Lee, "Computer vision-based human body segmentation and posture estimation," *IEEE Trans. Syst., Man, Cybern., Part A: Syst. and Humans*, vol. 39, no. 1, pp. 119–133, 2009.

[122] V. Powar, A. Jahagirdar, and S. Sirsikar, "Skin detection in YCbCr color space," *IJCA Proc. Int. Conf. Comput. Intell. (ICCIA2012)*, vol. iccia, no. 5, 2012, published by Found. Comput. Sci., New York, USA.

[123] M. Storring, H. J. Andersen, and E. Granum, "Skin colour detection under changing lighting conditions," in *Proc. 7th Symp. Intell. Robot. Syst.*, 1999, pp. 187–195.

[124] L. Bergasa, M. Mazo, A. Gardel, M. Sotelo, and L. Boquete, "Unsupervised and adaptive Gaussian skin-color model," *Image and Vision Computing*, vol. 18, no. 12, pp. 987–1003, 2000.

[125] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001, pp. I–511–I–518.

[126] J. Chung, P. Kannappan, C. Ng, and P. Sahoo, "Measures of distance between probability distributions," *Journal of Mathematical Analysis and Applications*, vol. 138, no. 1, pp. 280–292, 1989.

[127] Y. Kameda and M. Michihiko, "A human motion estimation method using 3-successive video frames," in *Proc. Conf. on Virtual Systems and Multimedia*, 1996, pp. 135–140. [Online]. Available: http://ci.nii.ac.jp/naid/10024346616/en/

[128] J. C. SanMiguel and S. Suja, "Skin detection by dual maximization of detectors agreement for video monitoring," *Pattern Recognit. Lett.*, vol. 34, no. 16, pp. 2102–2109, 2013.

भारतीय प्रौद्योगिकी संस्थान गुवाहाटी
Indian Institute of Technology Guwahati