FACE SEGMENTATION IN THERMAL IMAGES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MELİS ERYILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

FEBRUARY 2015

Approval of the thesis:

FACE SEGMENTATION IN THERMAL IMAGES

submitted by **MELİS ERYILMAZ** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. M. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences**   _____

Prof. Dr. Gönül Turhan Sayan
Head of Department, **Electrical and Electronics Eng.**   _____

Prof. Dr. Gözde Bozdağı Akar
Supervisor, **Electrical and Electronics Eng. Dept., METU**   _____

Examining Committee Members:

Prof. Dr. Kemal Leblebicioğlu
Electrical and Electronics Eng. Dept., METU   _____

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Eng. Dept., METU   _____

Assoc. Prof. Dr. İlkay Ulusoy
Electrical and Electronics Eng. Dept., METU   _____

Assist. Prof. Dr. Fatih Kamışlı
Electrical and Electronics Eng. Dept., METU   _____

Assist. Prof. Dr. Behçet Uğur Töreyin
Electrical and Electronics Eng. Dept., ÇANKAYA UNI.   _____

Date: 06/02/2015

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Melis ERYILMAZ

Signature   :

ABSTRACT

FACE SEGMENTATION IN THERMAL IMAGES

Eryılmaz, Melis

M.S., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Gözde Bozdağı Akar

February 2015, 87 pages

Automatic face segmentation is a key issue in many applications such as machine vision, coding, etc. Therefore, the accuracy of the segmentation algorithms results has a strong impact on the later stages. These algorithms should also be computationally efficient and robust against changing environments. The aim of this thesis is to analyze different approaches for face segmentation and compare them in terms of the robustness and computational efficiency. Four different face segmentation methods are chosen to be compared in the scope of this thesis. Experiments are performed on IRIS and Terravic databases. Implemented face segmentation methods are compared according to their classification performances and error rates.

Keywords: IR face segmentation, Bimodal Gaussian distribution model, morphological operations, thermal images, active contours, LSE curve fitting

ÖZ

TERMAL GÖRÜNTÜLERDE YÜZ SEGMANTASYONU

Eryılmaz, Melis

Yüksek Lisans, Elektrik Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Gözde Bozdağı Akar

Şubat 2015, 87 sayfa

Otomatik yüz segmentasyonu; bilgisayarla görü, kodlama vb. uygulamalarda anahtar rol görevini üstlenmektedir. Bu sebeple, segmentasyon algoritmalarının sonuçlarının, sonraki adımlara olan etkisi oldukça fazladır. Aynı zamanda, bu algoritmaların değişen koşullara dayanıklı, hesaplama açısından verimli algoritmalar olması önem arz etmektedir. Bu tezin amacı; farklı yüz segmentasyon metotlarını analiz etmek ve bu metotları dayanıklılık ve hesaplama açısından verimliliklerine göre karşılaştırmaktır. Bu tez kapsamında, dört farklı yüz segmentasyon metodunun uyarlanması ve bu yöntemlerin karşılaştırılması sunulmuştur. Deneyler IRIS ve Terravic veri tabanları için gerçekleştirilmiştir. Uyarlanan yüz segmentasyon metotları hata oranları ve sınıflandırma performanslarına göre karşılaştırılmıştır.

Anahtar Kelimeler: Kızılötesi yüz segmentasyonu, iki doruklu Gaussian dağılım modelleri, morfolojik işlemler, termal görüntüler, aktif konturlar, küçük kareler kestirim yöntemi ile eğri yerleştirilmesi

To my lovely grandfather, grandmother, my aunt and my mother

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

xiv

CHAPTER 1

INTRODUCTION

Face segmentation is an essential step for the face recognition systems since most of the face classification algorithms tend to only work with face images. Face segmentation efficiency is an important part of the real world face recognition applications.

Initially, face segmentation has been studied on the visible spectrum because of its low cost and wide range of usage. But the face segmentation in the visible spectrum is impractical because of the most challenging factors like; lighting variability, pose variations and noise [1], [2]. These factors significantly decrease the accuracy of the face segmentation results. To eliminate these negative effects and develop robust methods, IR images have become an alternative field of study where the face segmentation algorithms can be applied [3].

Thermal imagery may be obtained with different cameras which have different characteristics in terms of bandwidth usage. Infrared band is divided into different sub-bands such as; Near-Infrared, SWIR (Short-wavelength Infrared), MWIR (Mid-wavelength Infrared), LWIR (Low-wavelength Infrared) and Far-Infrared. IR face segmentation is generally using the LWIR region [4]. IRIS and Terravic databases are captured in LWIR band [5].

Thermal images made available the detection of the temperature differences easily. Therefore, temperature differences between the face and the body as well as the

differences between different parts of the face are detected more easily. But, this advantage of the thermal images is not enough to segment the face properly from the image. Temperature differences between the face or body parts, like cold nose and ears; warm clothes; neck and the shoulders are the main factors making the segmentation difficult in IR images. Contrary to the visible range images, face segmentation in thermal imaging is still an open area for improvement. Therefore, implementing the most widely used face segmentation methods used in thermal imaging, the efficiency and the robustness of the methods will be compared.

1.1   SCOPE AND OUTLINE OF THE THESIS

The scope of this thesis is to implement four face segmentation methods and then compare the experimental results of these methods on IRIS and Terravic databases. Firstly, face segmentation literature on the electro-optical and the thermal images is reviewed and important points for the IR face segmentation methods are defined in Chapter 2. Four IR face segmentation methods are selected to be implemented. All selected methods are analyzed, required mathematical models are derived and the methods are implemented on MATLAB platform except Pavlidis et al. [6] method; because the source code of this method is obtained from the internet [7]. All related algorithms and sub-methods for the methods are given in detail in Chapter 3. Then, the simulation results are analyzed individually and compared in Chapter 4 and the most effective method for face segmentation among all others is determined. In Chapter 5, the important features of the analyzed methods are summarized and the chapter is concluded presenting the future works.

CHAPTER 2

LITERATURE REVIEW

Face segmentation problem have long been studied both on electro-optical and thermal images. The overview of the existing methods in the literature is given below.

2.1   Methods Used In Face Segmentation Algorithms

Methods used in face segmentation can be classified into three main groups as; pixel based, edge based and the region based face segmentation methods.

2.1.1   Pixel Based Face Segmentation Methods

Color is an important feature to detect human faces. In color images, face can be segmented from the skin tone information of the input color image. The motivation behind this approach is that the human face has very consistent colors which are distinct from the colors of many other objects. There are so many types of the color spaces like; RGB, YCbCr, HSI (Hue, Saturation, Intensity), HSV (Hue, Saturation, Value), HSL (Hue, Saturation, Lightness), as well as many others. The most common color models are RGB, HSV and YCbCr.

RGB color model is sensitive to light. In comparison with other models like YCbCr or HSV, it has a disadvantage that it cannot separate precisely the chroma and the

intensity of a pixel. Therefore, it is difficult to distinguish skin-colored regions, so this color model is not widely used in skin or face detection algorithms.

HSV color model consists of Hue (H), Saturation (S) and the Value (V) attributes. Hue and saturation attributes can be used while segmenting the face according to this color model space.

YCbCr color model represents color as brightness and two color difference signals; therefore this color space is robust to the lightning variations.

RGB is not preferred on skin or face segmentation because of the changing color information with the lightning conditions and the difficulty of face segmentation on complex background which includes similar color values with the face. Rather than RGB; normalized RGB, YcbCr and HSV color spaces have been proposed for their robustness to the lightning changes in the environment.

Face segmentation requires a small amount of computation and can be done regardless of pose using color information. Skin color can be a unique and easy-to-find feature for color images. Experimental results show that the most important feature to detect skin-like regions is the chrominance. Although human face has so many variations dependent to the race, age or to the person, chrominance value is distributed over a small range. But for a thermal image, only the intensity values from 0 to 255 is used to classify the image. So, this may cause the wrong classifications; because there may be so many irrelevant regions that have similar intensity levels.

On the other hand, elimination of the shoulders and the neck is important in face segmentation process. But the neck and the shoulders is included in the image after using the color based segmentation methods.

Wang et al. [8], proposed a method for face detection in MPEG video. This method consists of three stages as chrominance, shape and frequency information are used

respectively. In this method, RGB color space is transformed to the YCbCr values to obtain the chrominance values. This algorithm is studied on macro block level which is a lower resolution version of the video frames. This is preferred to obtain the results as much as faster than to have as much as accurate. The main reason of the first stage is to find roughly the skin areas. The second stage mask images are used to eliminate false detected face regions or false alarms by first stage. And finally in the third stage, DCT coefficients are used to verify the face detection.

R. Hsu et al. [9] used extra clues to find the face in the image. In addition to the chrominance components of color image, information extracted from the luminance (face color may change via luminance) and some heuristics about face shape are used.

Sobottka et al. [10], used color and shape information while segmenting the skin regions. HSV color space is used to determine the skin regions according to the hue and saturation values.

Another method used the color information is Lin et al. [11]. In this method, face candidates are searched using the color information, locating possible face blocks and detecting the eye location. Normalized RGB and the HSV color spaces are used to reduce the lighting effects.

To minimize the lighting effects, Alabbasi et al. [12] used RGB, HSV and YCbCr color spaces together. It is evaluated that this approach is more robust to the variations of lighting conditions.

Also in color based face segmentation, color attribute may change dynamically by lighting variations. So, for different environmental conditions, different face color values shall be formed. Therefore, basic color based methods cannot string along with the changing situations. To overcome this problem, training processes are applied to the color images. For example, the system is trained using Gaussian

models according to the skin color. Then, the possible skin locations are detected using this Gaussian model.

Single Gaussian model is a dynamic model to be able to adapt changing conditions. Each pixel in the image is tried to be fitted in a Gaussian model (μ, σ). If a pixel lies within the defined Gaussian model, then parameters of that distribution is updated according to the current pixel and this pixel is labeled as foreground.

$$\mu_t = (1 - \alpha) * \mu_{t-1} + \alpha * x_t \qquad (2.1)$$

$$\sigma_t^2 = (1 - \alpha) * \sigma_{t-1}^2 + \alpha * (x_t - \mu_t)^2 \qquad (2.2)$$

μ_t is the updated mean of the Gaussian, σ_t is the updated variance of the Gaussian model. α is the learning rate. It effects the contribution of the last pixel values to the model.

This method provides a dynamic way to separate face and non-face pixels depending on a Gaussian model. If a pixel is not matched with the defined Gaussian model for the face, it is labeled as background.

This method has so many advantages than other static methods. It can adapt smooth changes in the image adapting the model iteratively. This model can be a preferred for visible range images but for IR images this method may fail for some conditions. Thermal images reflect the radiation emitted from the objects. So, a human face has different regions which have different temperature values. Nose, cheeks or ears may appear as background on thermal images. In a similar way, warm clothes may look like the human face or skin.

So, Single Gaussian Model will be insufficient to solve this problem. Consequently, this method cannot work precisely with multimodal distributions since it has a single Gaussian distribution.

Single Gaussian model based training process may be applied to the various image data as color, intensity, chromaticity etc.

Thi B. Kwolek [13] manually segmented a set of images containing skin regions for generating a skin model. This model is generated from the chromacity value of the pixels. This model is constructed on the Single Gaussian model. The reason is indicated as to not deal with the Expectation Maximization algorithm which requires iterations.

Jalilian et al. [14] used YCbCr color space model to estimate the underlying Single Gaussian model. Then Bayes rule is used to classify the pixel by the defined Single Gaussian model. After this classification process, Skin Probability Image for the hand or the face is obtained.

Rather than Single Gaussian Model, Mixture of Gaussians includes different Gaussians within a model. In this model, each pixel is represented as K Gaussian distribution. In this model, the probability of the current pixel to be in that model is given in the Formula (2.3).

$$P(x|\theta) = \sum_{i=1}^{K} \omega_i * \aleph(\mu_i, \sigma_i^2) \qquad (2.3)$$

K is the number of distributions, ω_i is the weight for each distribution, μ_i is the mean and σ_i^2 is the variance of the distributions. $\aleph(\mu_i, \sigma_i^2)$ is the Normal distribution which can be defined as in the (2.4).

$$\aleph(\mu_i, \sigma_i^2) = \frac{1}{\sqrt[n]{2\pi}\sigma} e^{-\frac{1}{2\sigma_i^2}(x-\mu_i)^T(x-\mu_i)} \qquad (2.4)$$

Each pixel will have an effect to update the parameters of this function on the training part. Updated parameters will be ω, μ and σ^2. This update process is performed using training sets.

Conditional probability of data point x_j is shown in the equation (2.5) . The Expectation Maximization algorithm and the update of the parameters are described in the (2.6), (2.7) and (2.8).

$$p(\theta|x_j) = \frac{\omega_k p(x_j|\theta)}{\sum_{t=1}^{K} \omega_t p(x_j|t)} \tag{2.5}$$

$$\mu'_\theta = \frac{\sum_{j=1}^{N} p(\theta|x_j) x_j}{\sum_{j=1}^{N} p(\theta|x_j)} \tag{2.6}$$

$$\omega'_\theta = \frac{\sum_{j=1}^{N} p(\theta|x_j)}{N} \tag{2.7}$$

$$(\sigma'_\theta)^2 = \frac{\sum_{j=1}^{N} p(\theta|x_j) (x_j - \mu_j)(x_j - \mu_j)^T}{\sum_{j=1}^{N} p(\theta|x_j)} \tag{2.8}$$

N is the number of the pixels used in the training set. This algorithm stops when the absolute difference between the kth result of ω_θ and the $k+1^{th}$ result of ω_θ is smaller than a predefined ε value. After this training part, each pixel is going to be classified and the probability for each pixel is going to be calculated using the updated parameters. According to the calculated probability, the determination of the pixel classification is done.

Application of the mixture of Gaussian is suitable for thermal images. In a thermal image, foreground pixels and background pixels can be divided into two different groups in each. The reason is that the face may contain colder parts like nose and ears and the background may contain warmer parts like clothes. So, to overcome a wrong segmentation, face and background can be represented as the mixture of Gaussians. Warmer parts of the face and the colder parts of the face create the mixture of two Gaussians and the warmer parts of the background and the colder parts of the background form another mixture of two Gaussians. Each pixel can be tested if it is within one of these distributions.

As mentioned in the Single Gaussian model, mixture of Gaussian based training process may be applied to the various image data as color, intensity, chromaticity etc.

Pavlidis et al. [6] is based on the idea of describing the classes on Mixture of Gaussians distributions. In the method, face and background is defined as two bimodal Gaussian distribution. Dominant Gaussian distribution corresponds to the warm parts of the face, while the recessive Gaussian distribution corresponds to the cold parts of the face like nose or cheeks. Using the same approach, background is also modeled in the same way. Dominant Gaussian for the background model is cold part and the recessive distribution is the warm parts of the background like clothes. This modeling is based on a training process. Manually segmented skin and background images are trained using Expectation Maximization algorithm to obtain multiple distributions. After obtaining these models, the probability for each pixel is calculated via Bayesian theorem. According to the obtained results, pixel is labeled as foreground or background. This method is pose-invariant and also not very sensitive to the artifacts like glasses or hats. For these reasons the method is selected as one of the method to be analyzed in detail and to be implemented in Chapter 3.

Filipe, S., Alexandre, L.A., [5] method is based on the Pavlidis et al. [6] method. But in this method, to improve the existing probabilistic method, edge detection technics and some morphological operations are also applied to the images. The main need for morphological operations is to enhance the image to only obtain the face and eliminate the unnecessary parts. The feature improved in this method is also selected to be used in the implementation and therefore analyzed in detail in Chapter 3.

Another study is conducted by [15] using the mixture of Gaussian method. In this method, to overcome the shortcomings of different color spaces, three color spaces are used to form a Mixture of Gaussian model. Images in the form of the normalized RGB, YCbCr and HSV are trained to obtain the model. Therefore, rather than

having separate Gaussian model for each model independently, all color spaces models are merged into a one mixture of Gaussian model.

In addition to the color based methods, thresholding is also considered as a well-known segmentation method in the literature. Substantially, thresholding method uses the intensity information which is related with the color information. Therefore, thresholding methods may be included in color-based face segmentation methods as well.

An image may be segmented using a threshold value. If a 'T' threshold is chosen to separate two regions, the decision mechanism will be as (2.9):

$$g(x,y) = \begin{cases} 0 & f(x,y) < T \\ 1 & f(x,y) \geq T \end{cases} \tag{2.9}$$

This mechanism is for bimodal oriented images. But if the image has different threshold values for one region, then Adaptive Threshold mechanism is needed. Adaptive thresholding steps are as defined below.

- Divide the image into sub-images

- Select an initial estimate for threshold value 'T0'

- Segment the image using 'T0' value. After this segmentation process, the image will be divided into two regions as S1 and S2.

- Compute the average values for the S1 and S2 regions as m1 and m2.

- Compute the new threshold value as given in (2.10):

$$T' = \frac{(m1 + m2)}{2} \tag{2.10}$$

Repeat second and the fourth steps until difference in T' in successive iterations is smaller than a predefined parameter T0.

Thresholding is a fast and easy to implement method; but static thresholding is not suitable for changing environments.

If the image contains many regions with changing values, an adaptation is needed. However, adaptive threshold value may result with false classification of the pixels. Even if it is an adaptive value, it is not enough to distinguish pixels precisely. Objects may have holes, extraneous pixels etc. These factors may affect segmentation results negatively.

In thresholding, only the intensity value is considered. There is no guarantee that the pixels identified by the thresholding process are contiguous. Error rates get worse as the noise gets worse, simply because with the existence of the noise, the pixel will not represent the real value.

Marius et al. [16] used thresholding method to segment the face from the image. In this method thresholding value is detected for the Cb and Cr values.

Gasparini et al. [17] analyzed six (6) different color spaces (RGB, HSV1, HSI, normalized rgb, YCbCr, HSV2). Genetic algorithms and color space thresholding is applied in this method to obtain the proper skin maps.

Kumar et al. [18] transformed 3D RGB color space to the 2D coordinate system by a color triangle. All R, G, B values are described by a value between the [0,255] and a degree in a color triangle. Color centroid of each pixel is calculated according to this triangle. The corresponding value is placed into a color centroid hexagon region distribution model. According to this model, thresholding values for the face regions are determined and the image segmented according to these thresholding values.

2.1.2 Edge Based Face Segmentation Methods

In face segmentation methods based on edge information, the main objective is to find the contrast between the face and the background. A jump in intensity from one

11

pixel to another corresponds to an edge in the image. There are many ways to perform the edge detection. However, it may be grouped into two categories, as gradient and Laplacian. The gradient method detects the edges by looking for the maximum and minimum in the first derivative of the image. The Laplacian method searches for the zero crossings in the second derivative of the image to find edges. Sobel and the Canny edge detection methods are the most relied operators.

Edge detection is one of the important attributes for the face segmentation methods. The accuracy of the edge detection has a remarkable effect on the segmentation results.

Cho *et* al. [19] proposed a method based on morphological operations. In this method, the first purpose is to enhance the image quality to obtain significant results. The image noise is enhanced using median filters. Then, a linear transformation is applied to increase the image contrast. Edges are detected using the Sobel edge detection method. To enhance the edge results, morphological operations are applied. On the next step, largest contour which may be the best candidate for the face contour is full filled. To obtain only the face part of the image, the difference between the full-filled image and the edge image is found. To obtain the segmented image, difference image is multiplied with the original one.

In this method, face contour must be a closed contour to have valid results. But this is not possible for all cases. If the face contour is not closed, the method will not give the expected results. On the other hand, this method works well for the shoulder elimination from the image. Therefore, as being one of the selected method details will be explained in Chapter 3.

Another shape based face segmentation method is used by Gyaourova *et* al. [20]. Gyaourova *et* al [20] applied an elliptical mask method for face segmentation that is first applied by Selinger *et* al [21]. In this method, an elliptical mask is used to segment the face from the background. To do this, the image is subsampled by a

factor of 10 in each dimension and then an elliptical binary mask is created to fit into the face. Then the subsampled image and the mask are multiplied in bit-wise. So, the masked image is obtained after this operation.

Another method which is used the elliptical mask method is Segundo et al. [22]. Face segmentation is achieved using the combination of the edge detection, region clustering and shape analysis. After obtaining the homogeneous regions applying edge detection and region clustering methods, elliptical shape is matched with the face part of the image.

But, these methods will work only on the same size images. Also, to use this method effectively, faces must be frontal, centered and captured at the same distance.

Chan, T.F., Vese, L.A., [23] proposed a method to define the boundaries of the objects are not necessarily defined by the gradient. This is obtained by minimizing the given energy function. This method can work on two phases. Color and gray images are segmented with this method. This method is sensitive to the noise and hair. Its computational burden is very expensive. The estimation of the initial point is very important and critical while converging to the result. This method will always converge to a solution whether it is the expected result or not. However the power of the method in determining the boundaries of objects is selected as one of the attributes to be investigated in detail in Chapter 3.

Filipe, S., Alexandre, L.A., [24] developed another face segmentation method to improve the deficiencies of the Pavlidis et al. [6] method. This method may be considered as the combination of the intensity based (for thermal images intensity is considered in color based segmentation methods) and shape based methods. This method is improved using LSE algorithms to find the curvature of the chin or the shoulders. After this detection, using active contour method, the face is segmented from the image. The details of this method will be explained in Chapter 3.

F.Shih and C.Chuang [25] used raster scan to find extreme points of the face all over the image from top to bottom and left to right.

2.1.3   Region Based Face Segmentation Methods

2.1.3.1   Face Segmentation Methods Based On Eigenfaces

This method is based on the eigenvectors and eigenvalues of the obtained vectors for face detection. Eigenvectors are formed using existing images in the database. These eigenvectors are somehow the set of features. Using eigenvectors, variations between different faces may be characterized. Each face image can be represented by the linear combination of the eigenvectors.

This method may fail while multiple faces exist in the image. This method may be intolerant to the rotations and tilted faces if eigenspace does not contain such information. And also, in a face segmentation process, neck is not included in the segmented part. While using the eigenfaces, size of the eigenface is the determinant criteria to eliminate or include the neck from the image.

K.Wong et al. [26] used eigenfaces method while fitting the eye candidates. The fitness for each eye candidate is calculated projecting them to the eigenfaces.

Butakoff et al. [27] is based on the eigenfaces method defined in [28]; but the main motivation of this method is to reduce the execution time while parallelizing the training part.

2.1.3.2   Face Segmentation Methods Based On Neural Network

The structure of the Neural Networks is similar to the real neurons. There is input layer, hidden layer and output layer. There are number of components on each layer

as defined by the system. Each neuron has a weight and the output of this neuron is obtained by the multiplication of the input and the defined weight. All inputs are summed together and passed through an activation filter. The output may be the input for the next layer or may be the output for the system. Different characteristics may be trained on the neural network. For example only the red chrominance value is trained in one of the face segmentation method in the literature.

The complexity of the network is the disadvantage for this method. The processing time may take long time depending on the numbers of the neurons and the neural network to be trained. To improve the efficiency of the system, neural networks need a lot of training samples. So, the processing effort of this method is very high.

J.Dargham et al. [29] used neural networks method. In neural networks, neural may be trained with different feature; but in this method only the red chrominance value of the normalized RGB color scheme is trained.

Feng [30] is also used neural network method to segment the face from the image. Color-ratio image is found related to the existence of the red chrominance value in the pixels. Eyes and eyesbrows have lighter and the lips have darker gray levels in the color-ratio image. In the light of this information, neural network structure is formed and the face is segmented from the image.

2.1.3.3 Face Segmentation Methods Based On Genetic Algorithms

This method provides to solve the problems with the multiple solutions. But there is no guarantee that the method will find generally the global optimum solution. This problem may occur when the populations have so many subjects. This method may take long time generally. GA requires long training times and training process directly affects the efficiency of the results.

K.Wong et al. [26] selected to study on Genetic Algorithms to segment the face. In this method, pair of eye candidates is selected among various possible blocks by means of genetic algorithms. Fitting process is calculated using Eigenfaces approach. After obtaining the fitness values, candidates which have higher values are selected for the verification process.

2.1.3.4   Face Segmentation Methods Based On Voronoi Diagram

This method relies on the symmetry while extracting the face boundary from the binary facial image. Using the edges of the image, Delaunay triangles (or Voronoi diagram) are formed. In this method all triangles are examined geometrically. The main purpose of these triangles is to link broken edges on the image. Thus, the broken edges of the face will be abstracted from the background.

This method is difficult to implement for the complex backgrounds and also this method causes computational load. This method is unable to overcome the face rotations, face with bread, face wearing glasses, the joining of the eyebrows or hair to the eye. This is the reason why this method is not a suitable method for orientation-independent. Also, it is not suitable for segmentation on the images that have artifacts like hats or glasses or the mask.

Xiao, Y., and Yan, H., [31] used Voronoi Diagram for segmentation process. This method is efficient to find broken edges on the face and to link them; but this method may fail in a complex background image. Also, this method is sensitive to the rotations, glasses, beard and noise.

2.2   Face Segmentation On IR Images

Previous chapter includes the methods used in the thermal and electro-optical images for face segmentation purposes. It is observed that, the studies related to the

electro-optical images have a longer past than the studies conducted on the IR images. Thermal imaging is developed to avoid the lighting variations in the visible range imaging. But this transition is not enough to solve all problems on segmentation because thermal images have come with some other problems as well. For example; opacity of the glasses caused the difficulties on thermal face segmentation process. Also, thermographs are sensitive to the temperature changes. It is observed that, with the growing potential of thermal imaging studies, the need for proposals in thermal image face segmentation became more pronounced. To bridge the gap, some methods developed for electro-optical images are tried to be adapted to the IR images; but it is seen that electro-optical images and IR images have different features. For example, visible range images carry so many information to describe and distinguish face like; color, chrominance and the face features like the location of the eyes, mouth etc. On the other hand, IR images carry only the intensity value. The limitation of the training sets is also one of the constraints for the implementation of the methods. Also, artifacts like hat or glasses are mentioned as a general problem for image segmentation on IR images.

It is observed that the problems with the IR face segmentation may vary in different manners. It is mentioned that electro-optical images are sensitive to lighting variations. But it doesn't mean that the IR images are fully robust to changing conditions. For example; IR images are sensitive to the temperature changes. Therefore, it is important to define dynamic algorithms while segmenting the face on IR images.

A great majority of the methods included the neck as a part of the segmented face; but normally neck is not included in face segmentation concept. In electro-optical images, shoulder may be eliminated easily according to the color information; but it is not easy to eliminate the shoulders and the chin from the segmented part on IR images, because these areas may have similar temperature values with the face. So it is significant to eliminate the neck and the shoulders from the segmented part of the face to have an accurate segmentation result.

To eliminate the neck and the shoulder easily from the image elliptical mask method is applied. Thus, only the image part included in the mask is considered as face; but this method has dome deficiencies like sensitivity to the face rotations.

Execution time may be an important factor as well. Recognition part may include serious calculation processes; because of this reason, long execution times or high computational burden for face segmentation may affect the efficiency of the overall method.

To decrease long execution times and high computational burden, easier methods are developed to segment the face from the image. But the accuracy of these methods is not effective as the other methods [24].

Moreover, a robust face segmentation method shall be pose-invariant. It means that the face rotations may not affect the segmentation negatively. But it is observed that some methods like eigenfaces or elliptical mask method is only applicable to the frontal images. Therefore, robust segmentation method should be independent from the face rotations and poses.

Face segmentation may be more demanding than the face detection; because segmenting a face not only means to detect the location of the face; but also means to determine the shape of the face. So, determining the face shape in detail is important to pave the way for the face recognition processes. For this reason, selection of the accurate shape detection methods is important.

All face segmentation methods may not involve all of these characteristics inside; but it is important to understand the differences, advantages, disadvantages and deficiencies of the existing methods.

Face segmentation on IR images have different features and difficulties from the electro-optical image face segmentation. Considering all of these important points and limitations, four reasonable methods which are developed especially for face

segmentation on IR images are implemented and compared to observe the effectiveness of the results.

Neural networks need so much processing time and a lot of training samples to be trained. This can be time consuming and the neural network method can occupy the source for following steps of the method. Eigenfaces may be intolerant to the rotations. Thresholding methods are sensitive to the noise. Genetic Algorithms requires long training times and training process to obtain effective results. Voronoi diagram is sensitive to the artifacts like hat or glasses. Considering all advantages and disadvantages of the methods mentioned above, Pavlidis et al. [6] algorithm is preferred to be implemented because it covers mixture of Gaussians method which enables to classify the face and background pixels with a probabilistic approach. (MATLAB code for Pavlidis et al. [6] method is not coded manually, the source code is found from the internet [7]. All other methods are coded in MATLAB manually.)

Knowing that, Pavlidis et al. [6] used an iterative method to segment faces from the image. It is observed that this method may take long time because of the iteration process. However, Cho et al. [19] face segmentation method is based on the morphological operations and edge detection processes. It does not contain iterative or complex processes, so this method is so fast when compared with the Pavlidis et al. [6] method. Rather than segmenting images with complex method, it is observed that it is possible to obtain valid outputs with less complex methods. To be able to compare this method's results with other methods and compare their efficiency, this method is selected as one of the implemented methods.

LSE curve fitting and Active contour approach is used by Filipe et al. [24] to obtain more accurate face segmentation results. In face segmentation concept neck and shoulders are not contained in the segmented part of the image. But majority of the existing methods do not put so many effort to work on this detail. But this method put emphasis on this detail and included this feature in the method. Furthermore, this

method tried to increase the accuracy of the results using active contour method. Also, this method is resistant to the artifacts like glasses, hats etc. So, this method is observed as a new, dynamic and accurate method for face segmentation in thermal images. Therefore, this method is also selected as one of the implemented methods.

Last implemented method is Filipe et al. [32] face segmentation method which is based on the Pavlidis et al. [6] method. In addition to the bimodal Gaussian model, some contour detection and morphological operations are also applied to the images to improve the results of the existing probabilistic method. The main need for these improving methods is to obtain only the face and eliminate the unnecessary parts of the body or background from the segmented part. So, using these supplemental methods, it is possible to eliminate clothes which are considered as face pixel in Pavlidis et al. [6] method.

CHAPTER 3

IMPLEMENTED ALGORITHMS FOR FACE SEGMENTATION

3.1   Bimodal Gaussian Approach by Pavlidis et al. [6]

In this approach, face and background pixels are classified by bimodal Gaussian models. Human face has a wide temperature distribution; as cold and hot parts. Parts of the face that are rich in vasculature correspond to hot parts of the face. On the other hand, cold parts consist of less vasculature, so the temperature of these parts is less than the other parts of the face. The background has also the same description; walls are considered as the cold objects, while clothing of the human can be considered as the hot objects. This structure leads to represent the distribution of the face as bimodal distribution. Hot parts of the face forms the dominant Gaussian distribution while the cold parts like nose or cheeks forms the recessive distribution. Similarly, same methodology is applied for the background pixels. Cold parts like walls form the dominant Gaussian distribution and the clothes form the recessive Gaussian distribution. The parameters like; ω, μ and σ for these Bimodal Gaussian distributions are obtained by EM algorithm using manually segmented skin and the background pixels.

The obtained bimodal Gaussian distribution results for the face and background are given in the Figure 3.1. As shown in the Figure 3.1 , dash lines represent the estimated background and estimated skin distributions. Continuous lines represent

the current background and skin pixels distributions obtained from the manually segmented training sets.

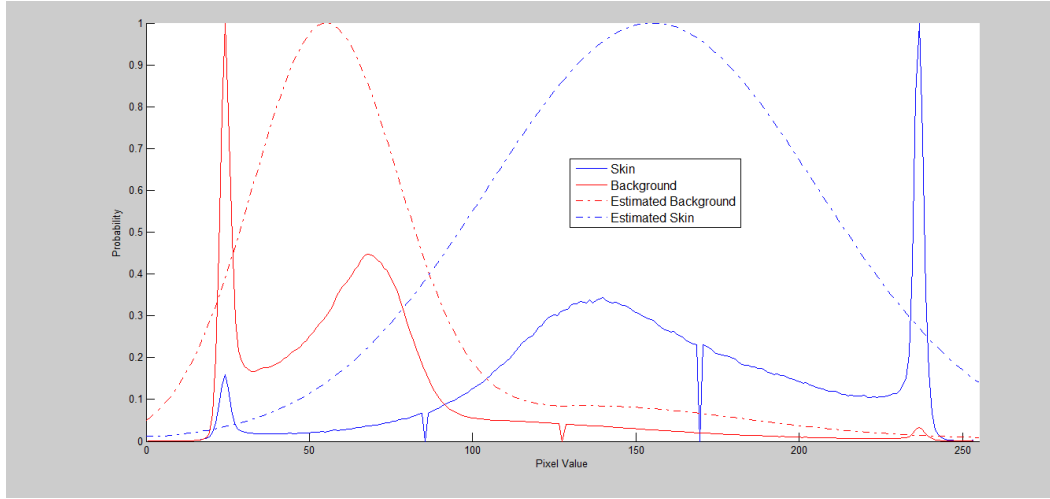This process is essential to obtain Gaussian models to be used later to classify each pixel.

Figure 3.1: Bimodal Distributions of Skin and Background

3.1.1 Pixel Classification

The main focus of this method is to estimate the classification probability for a given pixel. So, if it is assumed that the parameter of interest is θ, $p^{(t)}(\theta|x_t)$ is the unknown parameter that is trying to be solved. This parameter is also called "posterior distribution". The Bayes equation is given in the Formula (3.1).

$$p^{(t)}(\theta|x_t) = \frac{p^t(\theta)p(x_t|\theta)}{p(x)} \tag{3.1}$$

The parameter $p^{(t)}(\theta)$, represents the "prior distribution" that represents the probability that if a pixel is skin or background without any data at time t. The formula for prior distribution is given in (3.2).

22

$$p^{(t)}(\theta) = \begin{cases} p^{(t)}(s) & \text{when } \theta = s \\ p^{(t)}(b) = 1 - p^{(t)}(s) & \text{when } \theta = b \end{cases} \tag{3.2}$$

$p(x_t|\theta)$ represents the "conditional distribution" for the incoming pixel value x_t, which depends on whether the particular pixel is skin (θ=s) or background (θ=b) at time t. The related equation is given in the Formula (3.3).

$$p(x_t|\theta)$$
$$= \begin{cases} p(x_t|s): \omega_s^{(t)}\aleph(\mu_{s1}^{(t)}, \sigma_{s1}^{2(t)}) + (1 - \omega_s^{(t)})\aleph(\mu_{s2}^{(t)}, \sigma_{s2}^{2(t)}), & \text{when } \theta = s \\ p(x_t|b): \omega_b^{(t)}\aleph(\mu_{b1}^{(t)}, \sigma_{b1}^{2(t)}) + (1 - \omega_b^{(t)})\aleph(\mu_{b2}^{(t)}, \sigma_{b2}^{2(t)}), & \text{when } \theta = b \end{cases}$$

$$\tag{3.3}$$

As it can be seen from the likelihood equations, there are unknown parameters like ω, μ and σ. These parameters will be initialized and updated according to the "*Expectation Maximization*" algorithm. Details of this algorithm will be given later.

In the Bayes theorem another parameter to be calculated is $p(x)$. So $p(x)$ can be written in details as in the given in (3.4).

$$p(x_t) = p^{(t)}(s)p(x_t|s) + p^{(t)}(b)p(x_t|b) \tag{3.4}$$

If $p(x_t|\theta)$ and $p^t(\theta)$ are known, it is possible to find $p(\theta|x_t)$. So, while beginning the classification from the first pixel x_1, $p^1(\theta)$ prior distribution information is assumed as ½ because there is no prior information about the distribution.

$$p^{(1)}(s) = p^{(1)}(b) = \frac{1}{2} \tag{3.5}$$

Next parameter is $p(x_1|\theta)$ conditional distribution. As mentioned before there are unknown parameters like; ω, μ and σ. So these unknown parameters have to be calculated before the pixel classification part. These parameters are calculated in the training part of the segmentation using Expectation Maximization algorithm.

$$p^{(t+1)}(\theta) = \begin{cases} p^{(t+1)}(s) = p^{(t)}(s|x_t), & when\ \theta = s \\ p^{(t+1)}(b) = p^{(t)}(b|x_t) = 1 - p^{(t+1)}(s), & when\ \theta = b \end{cases} \qquad (3.6)$$

After obtaining the posterior distribution for each pixel individually, this posterior distribution is used to provide the prior distribution for the next pixel as given in (3.6).

3.1.2   Expectation Maximization Algorithm

Bimodal Gaussian distributions for face and background pixels need to be formed using this algorithm before the pixel classification part. The calculation of the conditional distribution parameters for background (or face) begins with the manual segmentation of the N facial images off-line. Manually segmented inputs should include all possible background (or face) parts (for example, clothing) to maintain accurate results. N_s is the number of segmented skin pixels, while N_b is the number of segmented background pixels in N facial frames. It is assumed that background (or face) pixels are sampled from a mixture of two Normal distributions as in the Formula (3.7).

$$p(x_t|b) = \sum_{i=1}^{2} \omega_{b_i} \aleph\left(\mu_{b_i}, \sigma_{b_i}{}^2\right) \qquad (3.7)$$

Expectation maximization algorithm begins with the calculation of a $z_{ij}^{(k)}$ parameter. This parameter will be used to update ω, μ and σ parameters.

Before the calculation, it is important to mention that ω, μ and σ parameters initially have some crude estimates. These values are provided initially by a ".mat" file and ω parameters is defined in the source code and it can be changed manually.

For k=0,1,…; i=1,2; j=1,2,…, N_s

$$z_{ij}^{(k)} = \frac{\omega_{b_i}^{(k)} (\sigma_{b_i}^{(k)})^{-1} \exp\left\{\dfrac{1}{2\left(\sigma_{b_i}^{(k)}\right)^2}(x_j - \mu_{b_i}^{(k)})^2\right\}}{\sum_{t=1}^{2} \omega_{b_t}^{(k)} (\sigma_{b_t}^{(k)})^{-1} \exp\left\{\dfrac{1}{2\left(\sigma_{b_t}^{(k)}\right)^2}(x_j - \mu_{b_t}^{(k)})^2\right\}} \tag{3.8}$$

$$\omega_{b_i}^{(k+1)} = \frac{\sum_{j=1}^{N_b} z_{ij}^{(k)}}{N_b} \tag{3.9}$$

$$\mu_{b_i}^{(k+1)} = \frac{\sum_{j=1}^{N_b} z_{ij}^{(k)} x_j}{N_b \omega_{b_i}^{(k+1)}} \tag{3.10}$$

$$(\sigma_{b_i}^{(k+1)})^2 = \frac{\sum_{j=1}^{N_b} z_{ij}^{(k)} (x_j - \mu_{b_i}^{(k+1)})^2}{N_b \omega_{b_i}^{(k+1)}} \tag{3.11}$$

This algorithm keep running till the absolute difference between the kth result of ω_{si} and the $k{+}1^{th}$ result of ω_{si} is smaller than a $\mathcal{E}$ value. $\mathcal{E}$ value can be determined as a very small value to obtain accurate weight values. Using this algorithm, all unknown parameters are calculated for skin and background bimodal Gaussian distributions.

3.1.3 Results

After calculating the posterior distribution for the given pixel, decision part is the next step. Using the posterior distribution, it is possible to find if data point x_1 is a skin or background pixel. H_0 is the value that is assumed that x_1 is a skin pixel; H_1 is the value that is assumed that x_1 is a background pixel. After calculating these two different assumptions, their values are compared and the one with the higher probability is chosen as in (3.12).

$$\begin{cases} Accept\ H_0\ if: p^{(t)}(x_t|s) > p^{(t)}(x_t|b) \Leftrightarrow p^{(t)}(x_t|s) > 0.5 \\ Accept\ H_1\ if: p^{(t)}(x_t|b) > p^{(t)}(x_t|s) \Leftrightarrow p^{(t)}(x_t|b) > 0.5 \end{cases} \tag{3.12}$$

The algorithm is applied to the different databases. Terravic database application results are as shown in Figure 3.2, Figure 3.3 and Figure 3.4.
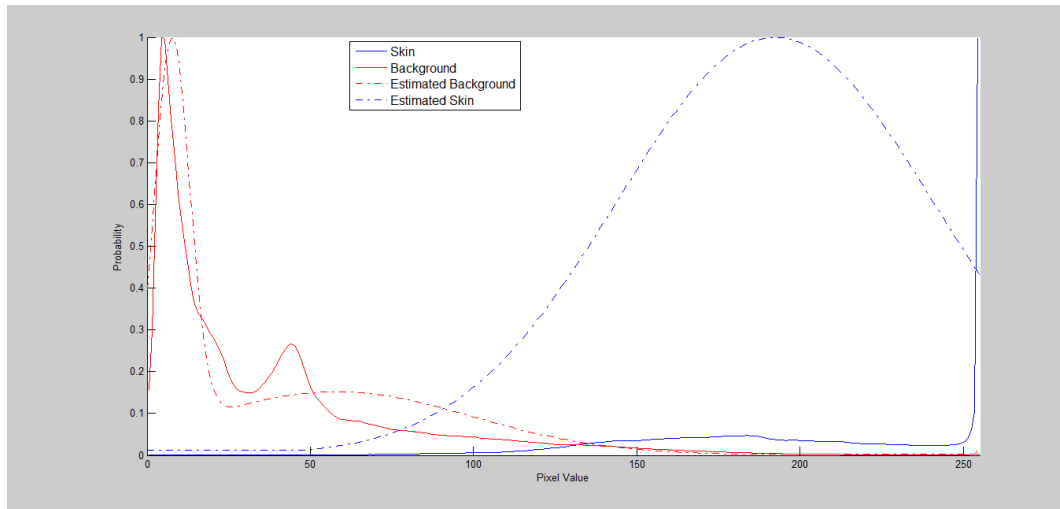

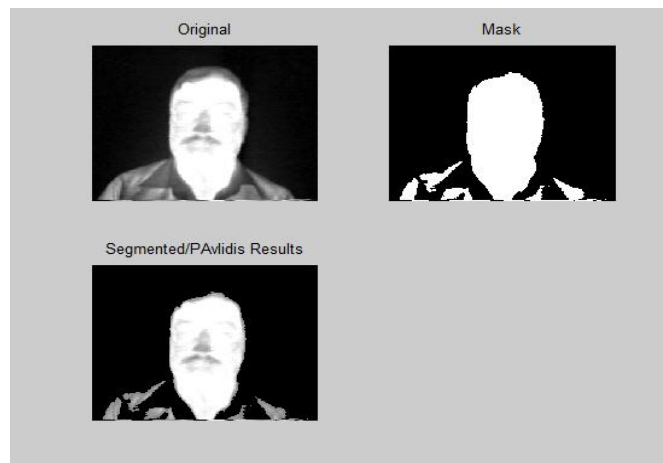Figure 3.2: Bimodal Gaussian Distribution for Terravic Images

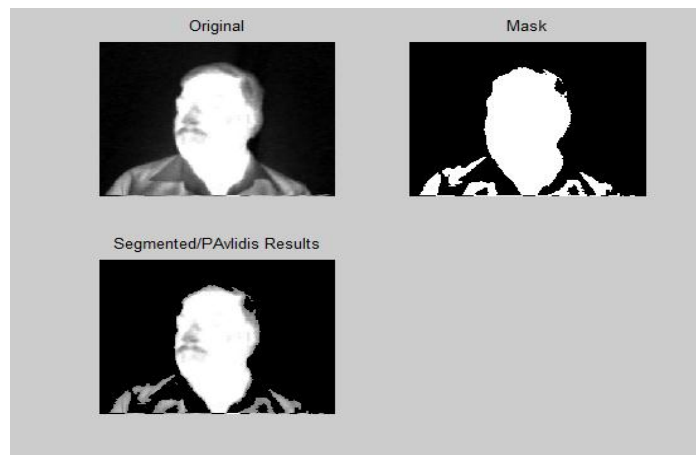Figure 3.3: Terravic Image Face Segmentation Result

Figure 3.4: Terravic Image Face Segmentation for a Different Pose

Results are not pose-variant. From the front and side poses the eyes and nose parts are considered as skin. But hair and some part of the cloth are considered as background. But some parts of cloth are also considered as skin as it can be seen from the results. Also, the moustache is also included in the face part. Algorithm has some deficiencies open for the future growth.

Bimodal Gaussian distributions for the IRIS database are given in the Figure 3.5 and Figure 3.6.
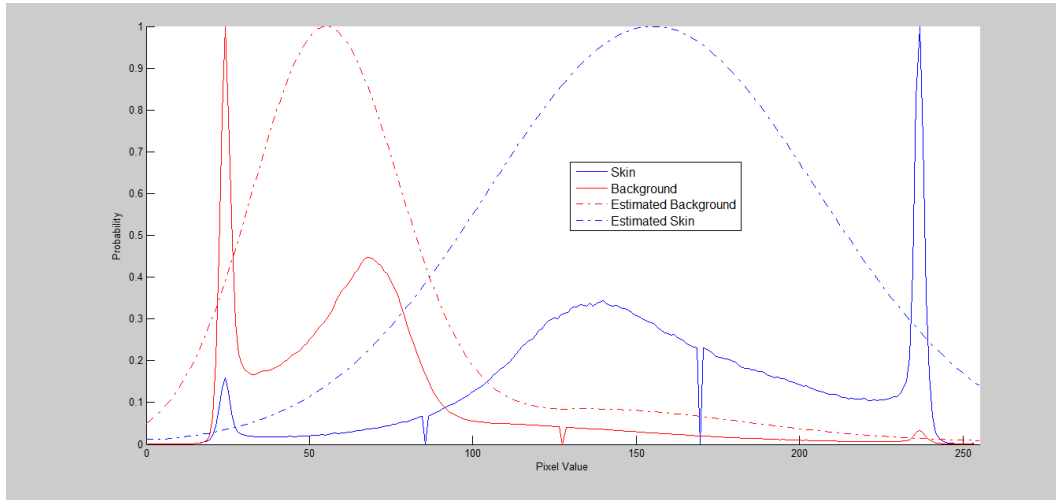
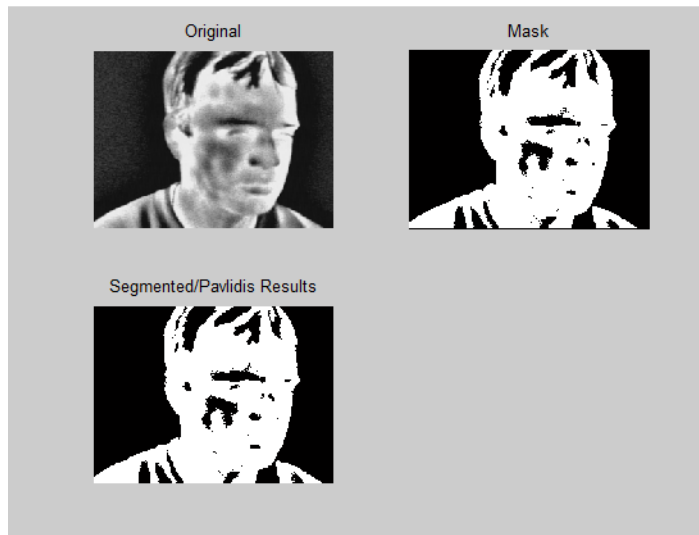Figure 3.5: Bimodal Gaussian Distribution for IRIS images

Figure 3.6: IRIS Image Face Segmentation Result

3.2 Closed Contour Approach by Cho et al. [19]

Cho et al. [19] approached the IR face segmentation in a different way. Morphological operators, image enhancement technics are the basics for this method.

28

In this approach, the image is removed from the salt and pepper noise by using the median filters. After this, to obtain significant results from the segmentation, image contrast is increased. A linear transformation is applied to the image for this purpose.

Then, edges are found using the Sobel edge detection method. But with this method, some disused edges are also included in the image. So, to remove these artifacts, some morphological operations are applied.

3.2.1   Morphological Operations

In this approach, to remove the artifacts, "bwareaopen" operation is applied. This operation erased objects smaller than defined pixel size. To connect the broken edges in all directions, "imclose" operation is applied using four different line structural elements. Length of the line structural elements is 5 and rotations are defined as 0, 45, 90 and 135 respectively. So, disconnected parts connected with the dilation operation and dilated parts returned to their original size with the erosion operation. If the image is considered as a matrix I with size NxM, and the defined structural element is assumed as B, the erosion and dilation operation can be cited as in Formula (3.13) and (3.14) respectively.

$$(I \ominus B)(x,y) = min_{(s,t)\in B}\{I(x+s,y+t\} \qquad (3.13)$$
$$(I \oplus B)(x,y) = max_{(s,t)\in B}\{I(x-s,y-t\} \qquad (3.14)$$

3.2.2   Face Center Detection

After all of these enhancement processes, the center of the face needs to be defined according to the method. The largest contour is detected using the length parameter of the existing contours. The contour which has the biggest length value is considered as the face contour. The center of this contour is found from the

coordinate array of the contour. Maximum and minimum x and y coordinates are calculated and the mean of these coordinates are found as the center coordinate of the face. After detecting the face contour center, this contour is full-filled as defined in the method [19].

The contour detection technic may differ from the original paper in this implementation; because in this method the biggest contour is chosen as the face contour.

To only obtain the face part of the image, the difference between this full-filled image and the edge image is found and the difference image is obtained. Then this difference image is masked with the original one. Then, this image is again enhanced to improve the contrast and the face segmentation is accomplished for this method.

3.2.3   Results

The effect of the morphological operations and linear transformation may differ for different databases; because the environmental conditions that the databases are captured are different. Also, the face contour may not give the expected results for soma cases, so the method may fail in these conditions. On the other hand, this method works well for the shoulder elimination from the image. The result of the method can be seen in the Figure 3.7.

Figure 3.7: Cho et al. [19] Method for Face Segmentation

3.3 LSE Curve Fitting and Active Contour Approach by Filipe et al. [24]

3.3.1 Formation of Face Limits

In this approach, Least Squares Estimation method is used in order to eliminate the shoulder or the neck from the segmented image. In addition, active contour method is applied to obtain accurate segmentation results.

The main motivation of this method is to augment the accuracy of the face segmentation results. The first contribution of this method is the elimination of clothes or the neck from the segmented parts. As mentioned before, clothes may be look similar to the faces on the thermal images because the body warms them. To overcome this issue, an operator is created in this method that contains the face. This operator is called "Rectangular Region of Interest" (RROI). To obtain this region, vertical and horizontal image signatures are analyzed. If I is assumed as an NxM matrix which describes the image, the vertical and horizontal signature vectors are computed as in the (3.15) and (3.16). Signature vectors are 1D vectors. These

31

vectors contain the sum of intensity of pixels along the columns and rows individually. The indexes of the column and the row are m and n for NxM image.

It is assumed that vertical signature 1D vector is V, and horizontal 1D signature vector is H.

$$H(n) = \sum_{n=1}^{N} I(n, m) \tag{3.15}$$

$$V(n) = \sum_{m=1}^{M} I(n, m) \tag{3.16}$$

Initially, vertical signature is calculated summing up all the columns of the image in order in a 1D vector. As expected, the image has high frequency components because of the noise factor. High frequency components become evident in the 1D sum vector as shown in Figure 3.8. To eliminate the high frequency components, the vector is smoothed with the Gaussian as shown in the Figure 3.8. The intensity changes become distinctly visible from the Figure 3.8. These parts are the extrema points of the first derivative of the 1D sum vector. After taking the first derivative, the extrema points will be determined as left and right face limits.

The main purpose to find face limits is to avoid shoulders to exist in the segmented image. Horizontal signature is used to discard the hair in the same manner; but the Gaussian filter is different from the previous filter. Filter is as defined in the Figure 3.10, and the face limits are shown in Figure 3.11. Upper limit is calculated but the down face limit is not found through this calculation; because as mentioned before, this method involves sub-methods to eliminate shoulders and neck from the image. To do this, a curve fitting algorithm is going to be applied in the following steps. So, down limit of the face is going to be calculated from this curve fitting algorithm.

In this implementation, Gaussian function parameters used for blurring function may differ from the original code.
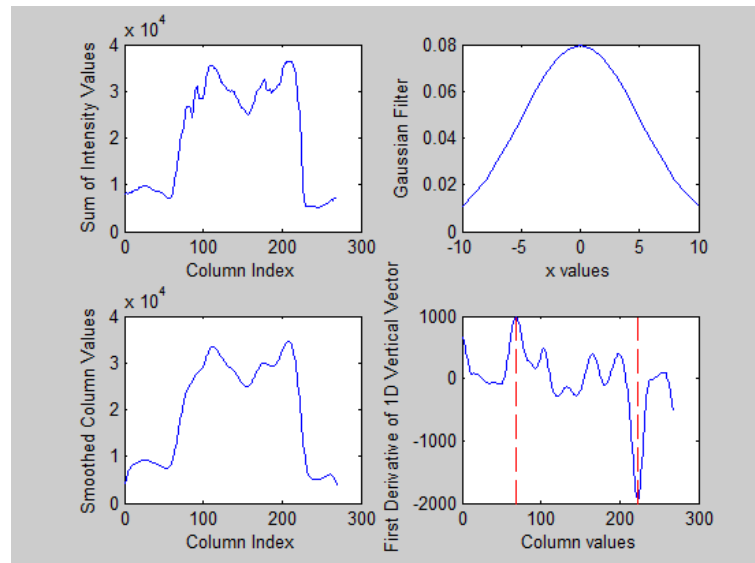
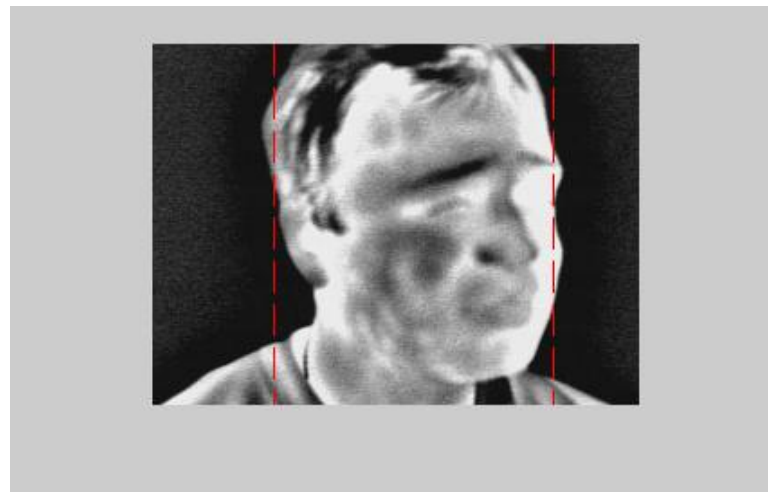Figure 3.8: Vertical 1D Signature and Gaussian Filters

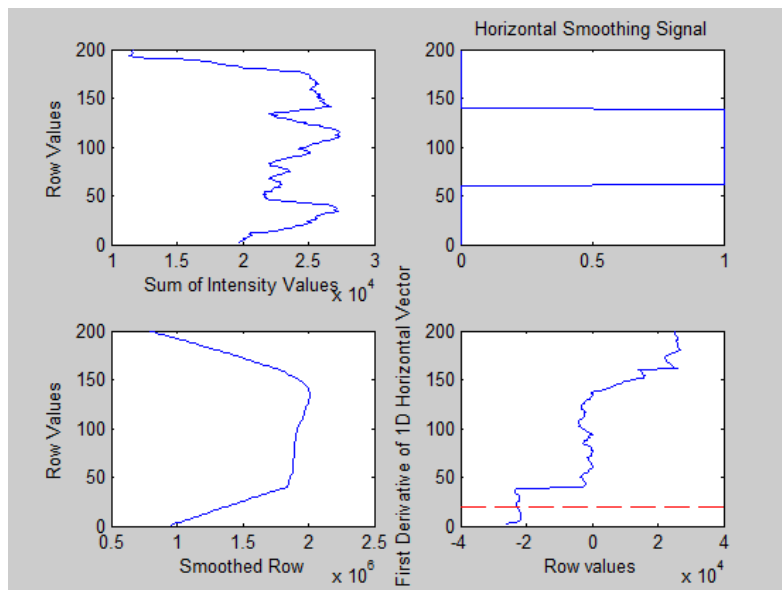Figure 3.9: Visualization of the right and left face limits

Figure 3.10 Horizontal 1D Signature and Gaussian Filters

Figure 3.11 Face limits of the face image

3.3.2 Curve Fitting Method with Least Squares Estimation

The elimination of the neck or the shoulders is one of the advantages of this method. This elimination process provides efficiency while finding the real face parts on the image.

LSE method will be applied only to the [2/3*R,R] part of the image. R is the number of rows. The reason of this limitation is based on the idea that the shoulder or chin will be located on the 2/3 part of the image. LSE method requires some pre-steps to obtain reasonable results. So, after cropping the image, linear reduction in the number of colors is applied to give prominence to the chin or shoulder. This reduction is also applied to decrease noise on the image. Next, to only obtain the main edges on the image, Gaussian blur is applied. Thus, the edges of the image is ready to be found by Canny edge detection method. After finding edges, LSE curve fitting method is applied. The LSE method requires input coordinates to calculate the best fitting curve to these points. The curve is fitted to a second order function $f(x) = ax^2 + bx + c$. The fitting curve is going to be found through the calculation of the unknown parameters a, b and c. The points are received from the Canny edge method as x and y coordinates. If each point is considered as x_i and y_i, least squares error for the dataset provided by Canny method can be found using the equation (3.17).

$$S = \sum_{i=1}^{M} \left(y_i - (ax_i{}^2 + bx_i + c) \right)^2 \qquad (3.17)$$

Since LSE method is based on the calculation of the minimum value of the square error, the equation given in (3.17) is going to be minimized. M is the number of coordinate pairs included in the dataset. S is the sum of the squared residuals. The residual is the difference between the actual value of the dependent and the value

predicted by the model. To find the minimum value of the sum of squares, the gradient is equalized to zero.

$$\frac{\partial S}{\partial a} = 0 \qquad (3.18)$$

$$\frac{\partial S}{\partial a} = 2 \sum_{i=1}^{M} (y_i - (ax_i^2 + bx_i + c))(-x_i^2) \qquad (3.19)$$

The equation (3.19) is equalized to zero, and the obtained equation is given in (3.20).

$$a \sum_{i=1}^{M} (x_i^4) + b \sum_{i=1}^{M} (x_i^3) + c \sum_{i=1}^{M} (x_i^2) = \sum_{i=1}^{M} (y_i x_i^2) \qquad (3.20)$$

Then, for each variable, the gradients are equalized to zero and the equations are obtained for each.

$$\frac{\partial S}{\partial b} = 0 \qquad (3.21)$$

$$\frac{\partial S}{\partial b} = 2 \sum_{i=1}^{M} (y_i - (ax_i^2 + bx_i + c))(-x_i) \qquad (3.22)$$

The equation (3.22) is equalized to zero, and the obtained equation is given in (3.23).

$$a \sum_{i=1}^{M} (x_i{}^3) + b \sum_{i=1}^{M} (x_i{}^2) + c \sum_{i=1}^{M} (x_i) = \sum_{i=1}^{M} (y_i x_i) \qquad (3.23)$$

And for the last variable c the equations found as given in (3.24).

$$\frac{\partial S}{\partial c} = 0 \qquad (3.24)$$

$$\frac{\partial S}{\partial c} = 2 \sum_{i=1}^{M} \left(y_i - (ax_i{}^2 + bx_i + c) \right)(1) \qquad (3.25)$$

The equation (3.25) is equalized to zero, and the obtained equation is given in (3.26).

$$a \sum_{i=1}^{M} (x_i{}^2) + b \sum_{i=1}^{M} (x_i) + cM = \sum_{i=1}^{M} (y_i) \qquad (3.26)$$

All of these equations can be transformed into a matrix form as shown in (3.27) and from this matrix, unknown values can be found from the division of matrices.

$$\begin{bmatrix} \sum_{i=1}^{M} (x_i{}^4) & \sum_{i=1}^{M} (x_i{}^3) & \sum_{i=1}^{M} (x_i{}^2) \\ \sum_{i=1}^{M} (x_i{}^3) & \sum_{i=1}^{M} (x_i{}^2) & \sum_{i=1}^{M} (x_i) \\ \sum_{i=1}^{M} (x_i{}^2) & \sum_{i=1}^{M} (x_i) & M \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{M} (y_i x_i{}^2) \\ \sum_{i=1}^{M} (y_i x_i) \\ \sum_{i=1}^{M} (y_i) \end{bmatrix} \qquad (3.27)$$

After finding a, b and c parameters, down face limit of the face is calculated according to the sign of the a parameter which gives information about the direction

37

of the curve. If a<0 then it means that the curve is fitted to the chin. If a>0 then it means that the curve is fitted to the shoulders. So down limit is calculated according this defined curves.

The steps of the LSE curve fitting method results are shown in Figure 3.12. This is an example to the curve fitting to the chin. Obtained results for the elimination of the shoulders and the neck are shown in the Figure 3.13 and Figure 3.14.

LSE curve fitting method used in this implementation may differ from the original paper; because in this implementation, LSE curve fitting is computed by mathematical model.
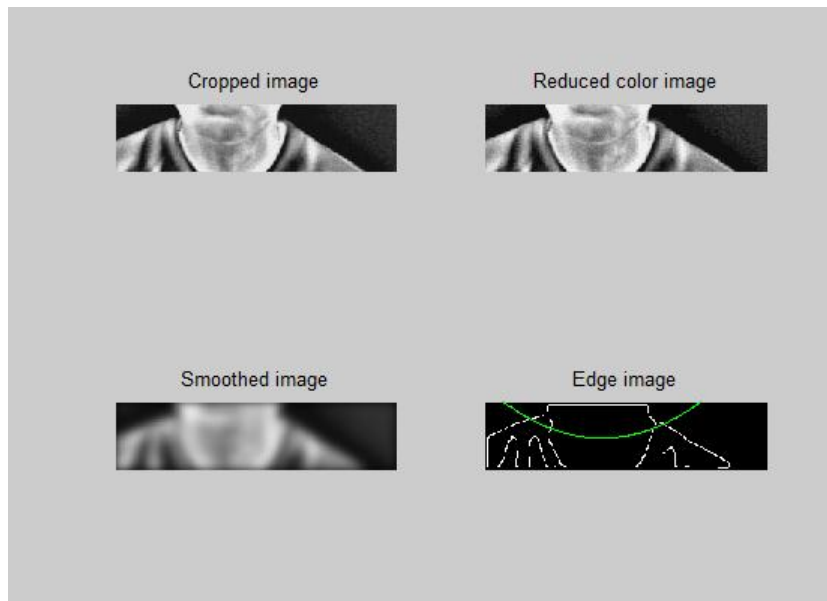

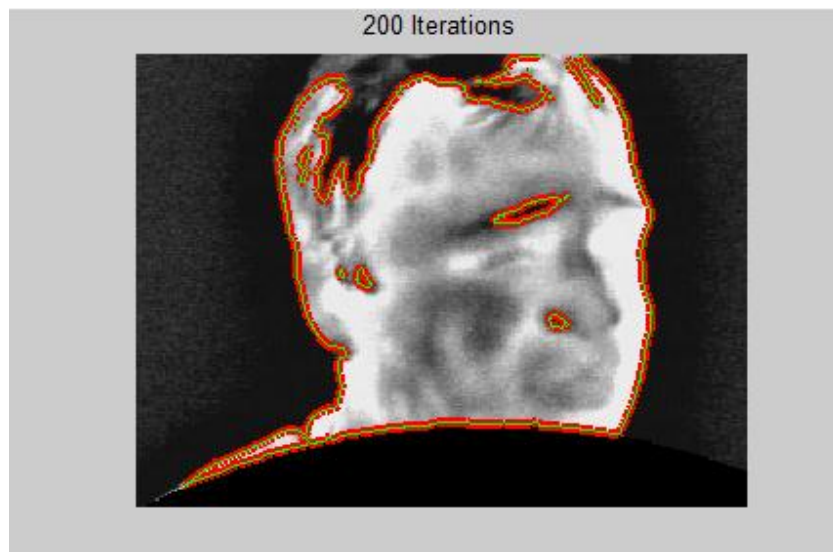Figure 3.12: LSE Curve Fitting Method Steps (fitted to the chin)

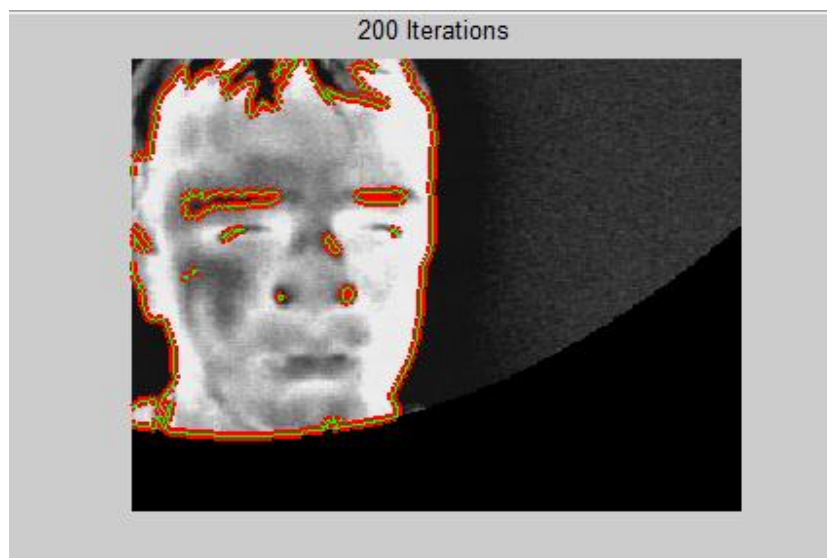Figure 3.13: An example for the shoulder elimination

Figure 3.14: Example for the neck elimination

3.3.3   Active Contours Method

Active contour method is used to segment the face from the image as accurate as possible. Chan and Vese [23] active contour method is used in this method.

Chan-Vese [23] active contour method is based on the idea to minimize the energy function. If the evolving curve is defined as C in Ω, ω is the region inside (C) an the outside (C) is defined as $\Omega - \omega$. It is assumed that u_0 is the entire image and the c_1 is interpreted as the mean value of everything inside the contour C and c_2 is interpreted as the mean value of everything outside the contour C. Regions and the contour are defined as in the Figure 3.15.
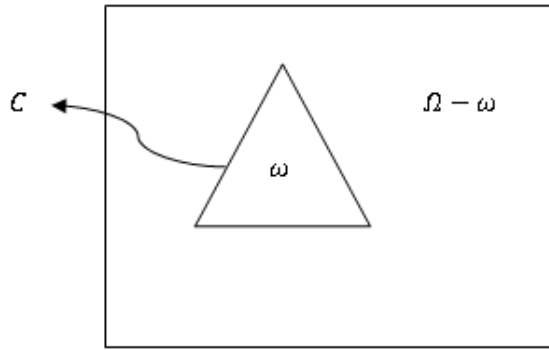

Figure 3.15 Contour and regions description

Following fitting term (3.28) is considered to explain general concept of the method.

$$
\begin{aligned}
F_1(C) + F_2(C) = &\int_{inside(C)} |u_0(x,y) - c_1|^2 dxdy \\
&+ \int_{outside(C)} |u_0(x,y) - c_2|^2 dxdy
\end{aligned}
\tag{3.28}
$$

In (3.28), $F_1(C)$ is defined as the force to shrink the contour and $F_2(C)$ is defined as the force to expand the contour.
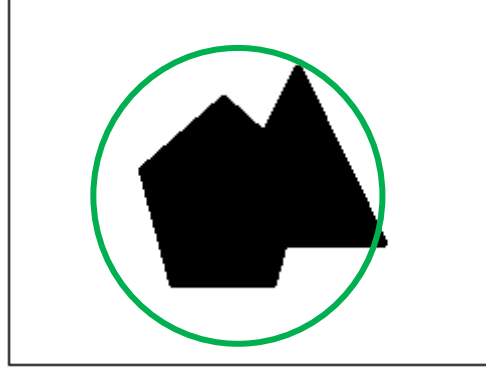

Figure 3.16 $F_1(C) > 0$, $F_2(C) \approx 0$, $Fitting > 0$

In Figure 3.16, white parts are considered as 1, black parts are considered as -1. From this equation, c_1 is considered as the mean of the inner contour and c_2 is considered as the mean of the outer contour. So, c_1 is approximately 0 (zero) and c_2 is equal to 1 (one). From (3.28), any pixel from the outside of C minus c_2 is equal to 0 (zero). So, $F_2(C)$ term is eliminated. Then, any pixel from the inside of C minus c_1 is equal to a positive value. So, $F_1(C)$ >0 and the contour shrinks to the inside to balance two forces. Another demonstration is given in Figure 3.17.
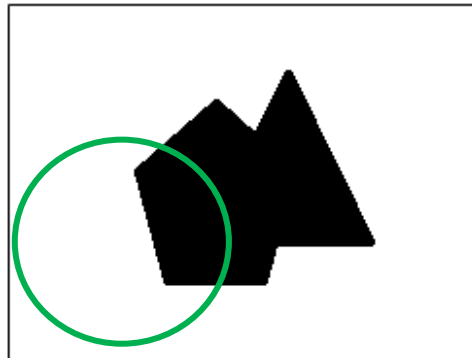
Figure 3.17 $F_1(C) > 0$, $F_2(C) > 0$, $Fitting > 0$

From the Figure 3.17, it is clear that c_1 and c_2 are values between 0 (zero) and 1 (one). So, $F_1(C) > 0$ and $F_2(C) > 0$. This means that the contour is going to grow internally and externally. The fitting finalizes when two of the force is equal to zero (0).

In Chan-Vese [23] active contour without edges method, the energy function is defined as in (3.29).

$$
\begin{aligned}
F(c_1, c_2, C) = {} & \mu. \, Length(C) + v. \, Area\big(inside(C)\big) \\
& + \lambda_1 \int\limits_{inside(C)} |u_0(x,y) - c_1|^2 dxdy \\
& + \lambda_2 \int\limits_{outside(C)} |u_0(x,y) - c_2|^2 dxdy
\end{aligned}
\tag{3.29}
$$

Generally λ_1 and λ_2 is equal to 1 (one) and v is equal to zero (0). After levet set formulation C which is represented by the zero level set of a Lipschitz function and considering the given parameters as defined before, the energy function is as defined in (3.30).

$$
\begin{aligned}
F(c_1, c_2, \phi) = {} & \mu \int_\Omega \delta\big(\phi(x,y)\big)|\nabla\phi(x,y)|dxdy + v\int_\Omega H(\phi(x,y))dxdy \\
& + \lambda_1 \int_\Omega (u_0(x,y) - c_1)^2 H(\phi(x,y))dxdy \\
& + \lambda_2 \int_\Omega (u_0(x,y) - c_2)^2 (1 - H(\phi(x,y)))dxdy
\end{aligned}
\tag{3.30}
$$

42

H(.) is the Heaviside function and $u_0(x, y)$ is the input image. To find the minimum value of the (3.30), its first derivative is taken and is equaled to zero.

Results for the active contour method are achieved with 200 iterations. The default value for μ is 0.2. Results are given as Figure 3.18, Figure 3.19 and Figure 3.20.
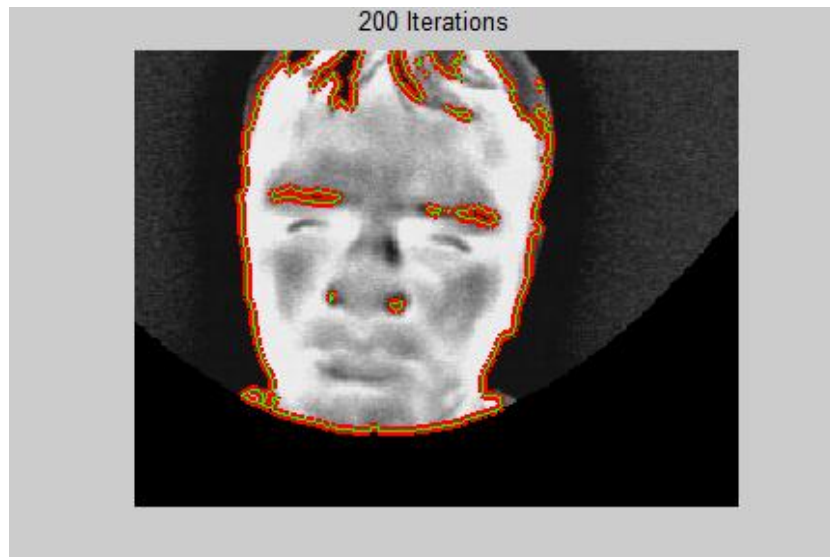

Figure 3.18 Chan-Vese [23] Active Contour Method-1

Figure 3.18 and Figure 3.19 are accurate segmentation results. However, in Figure 3.20 active contour segmentation result is not acceptable. The reason for this is the contrast difference of the pixels within the face and bad estimation of the initial point of the initialization contour.

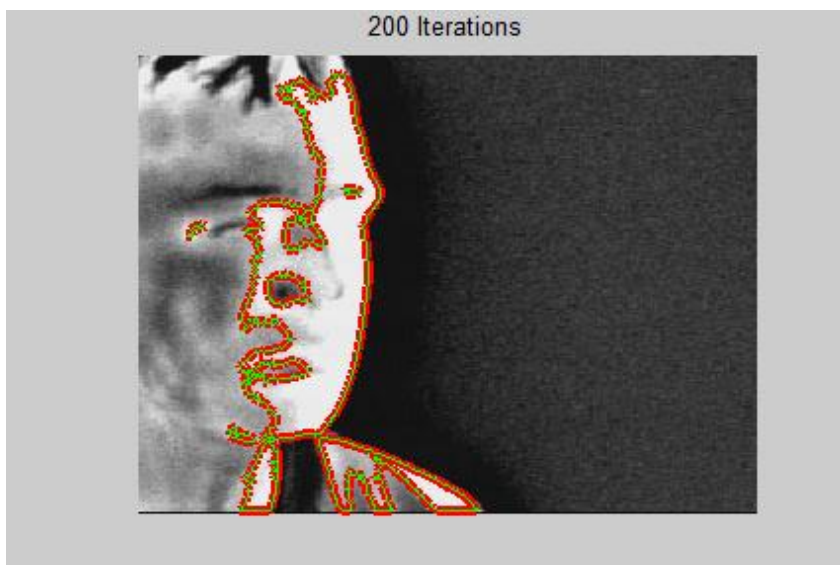Figure 3.19 Chan-Vese [23] Active Contour Method-2

Figure 3.20 Chan-Vese [23] Active Contour Method-3

3.3.4 Enhancement Methods for Face Pixel Identification

After finding potential face pixels using Active Contour method, to obtain accurate results, some enhancement technics have to be implemented. Even though active contour method is an efficient way to segment images, the segmented image needs to be examined totally.

First of all, the center of the face is identified using the extrema values of the signature vectors. Then small areas are removed from the image and edges are found using the Canny edge detection method. Dilation is also applied to enhance the edges. All pixels in the largest contour that contain the face center are going to be considered as face pixels. From the edges, the largest contour is found and considered as the face contour. Face contour is full-filled. But full-filling process of the biggest contour may cause wrong classification of the pixels. To overcome this problem and to reclassify the glasses pixels which are considered as face pixels before, the difference between the full-filled image and the segmented image with Chan Vese [23] Active Contour method is found. This difference image made possible to see filled glass pixels in the biggest face contour. So, to exclude the glass pixels from the full-filled image, the difference image is operated with an opening process. The main purpose of this opening operation is to eliminate all pixels smaller than the glass pixels and only obtain the wrongly classified glass pixels. Finally, OR operation is applied to this image with the full-filled image and the accurately segmented face image is obtained.

In this implementation face contour detection algorithm may differ from the original code; because in this implementation, all contours other than the face contour is eliminated using a self-developed algorithm. Face contour is assumed as the biggest contour. And also, structural elements used in the morphological operations may be different than the original code of the method.

3.3.5 Results

The results for the Filipe et al. [24] method are given in Figure 3.21. Final image is the last obtained result from the Figure 3.21.
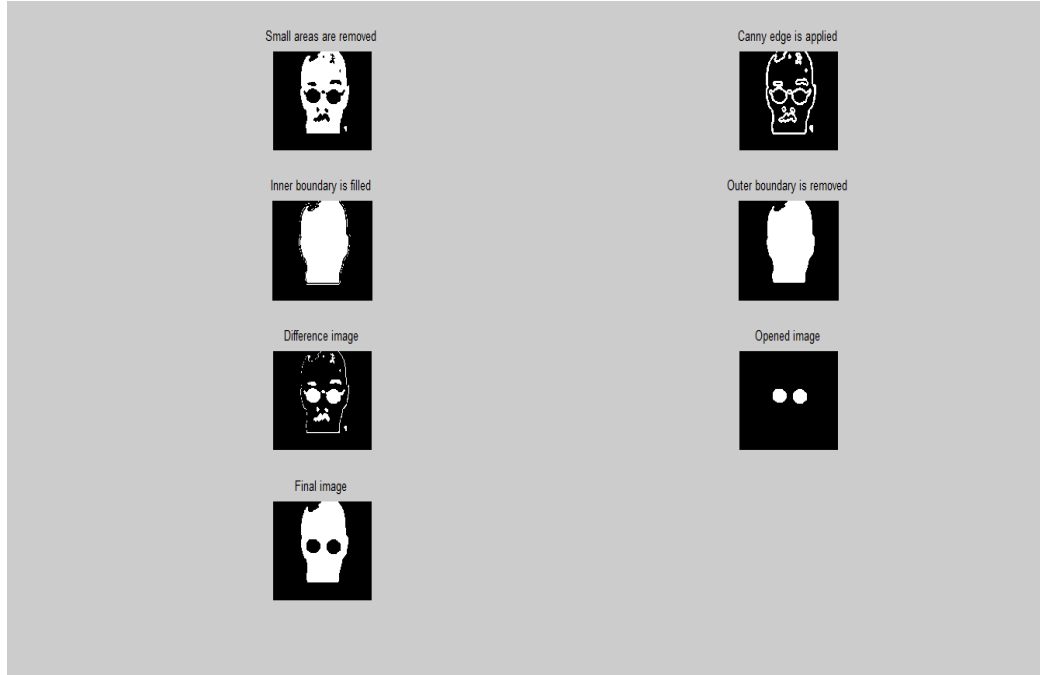

Figure 3.21: Enhancement process of the method

3.4 Signature and Contour Detection Approach by Filipe et al. [32]

3.4.1 Formation of Face Limits

The main motivation of this method is to improve the Gaussian model based face segmentation [6] results. In this approach, to remove the effect of clothing, vertical and horizontal pixel signatures are examined. To obtain vertical signature, pixels along the column are summed and for the horizontal signature pixels along the rows are summed as in the (3.15) and (3.16). Then, small areas in the image are filled

with 4x4 square structural element using dilation. After that, the proportion of each signature vector is found. The proportion is the value of the maximum signature value over the existing signature value. If the existing signature value is below the %25 of the maximum signature, then it is considered as background. So, this method is applied to both of the vertical and horizontal signatures. Doing this, major parts of the clothes are eliminated from the segmented image. The vertical elimination of the signatures which are smaller than %25 of the maximum signature caused some unexpected results as the removal of nose, some part of the cheeks and smaller face parts.

After the optimization of the background values using a threshold value, maximum values of the new signatures (for horizontal and vertical separately) are found. If there is more than one maximum value, the average of the values is calculated. Then, the center point of the face is found using these maximum values.

In the original paper the proportional value is defined as %20; but while implementing this method it is observed that the best results are obtained with %25 proportional value.

3.4.2 Morphological Operations

After marking the center point, 6x6 and 3x3 square structural elements are used to erode and dilate the image to remove some links between areas. Then, edges of the image are found using canny edge detection method. After finding the contours, 6x6 filter is used to enhance the contours obtained by the canny method.

Morphological operations used in the implemented code may differ from the original code. In the original paper, structural elements are chosen as 3x3 and 2x2 square structural elements. In this implementation, usages of the morphological operations are applied observing the output results.

3.4.3 Contour Detection

Boundaries on the image is found by *"bwboundaries"* method. This method gives all of the boundaries in the image. Till this step, image is processed using various methods as defined before. After this part, it is important to decide which boundary in the image is face contour. According to the method, the biggest boundary that has the center point inside is going to be accepted as the face boundary. So, to find the boundary which meets this criterion, the center point is assumed as the starting point while finding the face contour. Firstly, the point is shifted through the right. While shifting the point, x value (which means column) of the point is increased one by one while the y value (which means row) is stayed fixed. The meaning of the shifting operation is to find a non-zero value (an edge). It is known that a non-zero value which is located outer or inner of an object is included in a boundary. So, when a non-zero value is found, it is checked to find which boundary includes these coordinates. After finding the boundary which includes this pixel coordinates, this boundary is saved in the *'right_boundaries'* variable. All non-zero pixels are and their related boundaries are found scanning the x-axis through increasing y values. After finding right boundaries, same methodology is applied to the left side of the center point decreasing the x value and fixing the y value. All non-zero pixels and the boundaries which includes this pixels are also saved in the *'left_boundaries'* variable. As might be expected, *'right_boundaries'* and *'left_boundaries'* vectors contain unnecessary boundaries. We only need to have largest contour values. So, to eliminate unnecessary values, only the maximum values of these two sets are chosen. But, another control is also performed to check the difference between these maximum values. If the difference is much more than a threshold value, only the largest boundary (may be left or right boundary) is accepted as the face contour. The threshold value is assumed as the minimum of the *'right_boundaries'* and *'left_boundaries'* value. After all of these steps, the image needs to be cleared from all other unrelated boundaries. Elimination of the unnecessary boundaries process is completed as scanning all boundaries except the face contour. All coordinates of the

unnecessary boundaries are checked separately if there is another y point which has the same x value. If there is, all y values are recorded to a matrix with the related order. Then, maximum and minimum y values are found and then the distance between the maximum and the minimum values is calculated. Beginning with the minimum y coordinate and incrementing it one by one till the maximum y coordinate, all points between these coordinates are assigned as zero "0". So, all unnecessary boundaries are erased from the image.

Contour detection algorithm which is applied in this implementation may differ from the original implementation of this method.

3.4.4   Results

In the original implementation of this method, FSU and UND databases are used [32]. But, in this implementation, only the available databases are used (IRIS and Terravic databases). Two sample results are shown in the Figure 3.22 and Figure 3.23.

This method works well generally for the frontal images. In some non-frontal images, some part of the nose, cheeks may be deleted because of the face rotations. It means that this method may be defined as pose-dependent. Also, neck is included as a part of the face in the results.

As defined above, shoulders and some parts of the hairs are eliminated from the image (nose is also eliminated in non-frontal images as a disadvantage) using a threshold value for the signatures.
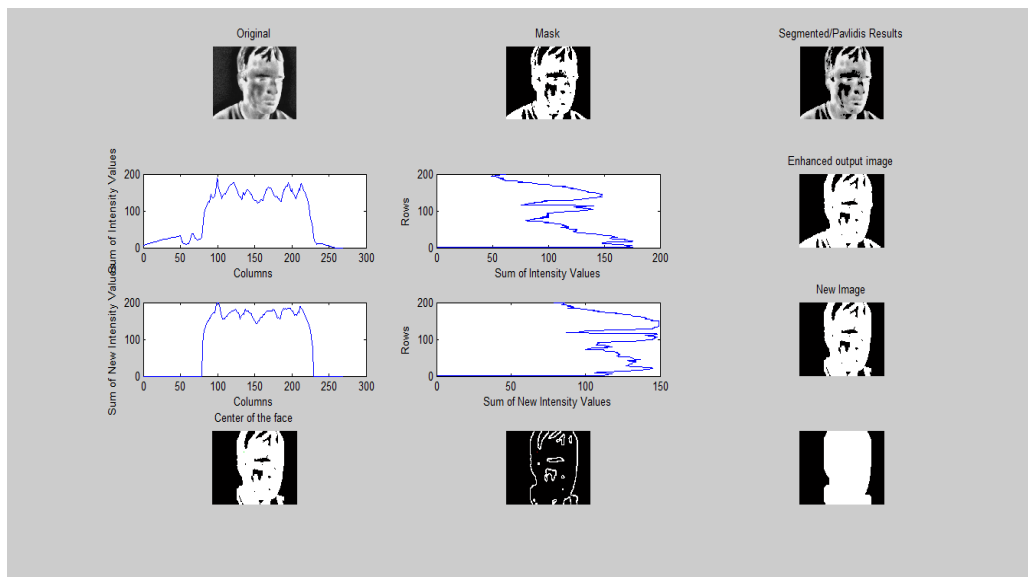
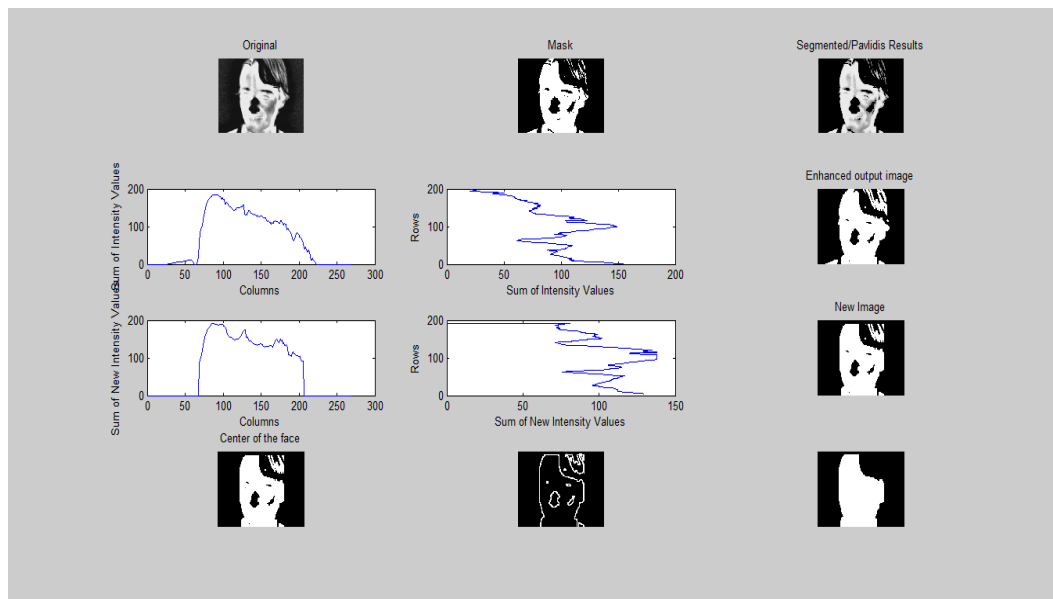Figure 3.22: Filipe et al. [32] Segmentation Result_1

Figure 3.23 Filipe et al. [32] Segmentation Result_2

CHAPTER 4

EXPERIMENTAL RESULTS

In this chapter four methods are compared according to their classification performances and error rates. Results are observed in two different databases for each method.

4.1   Datasets

Comparisons of these methods are performed on two datasets; 'IRIS Thermal/Visible Face Database' and 'Terravic Facial IR Database'. 'IRIS Thermal/Visible Face Database' contains 4228 images with different rotations, illuminations, expressions and artifacts like glasses or hat. 296 images are selected to be used in the methods. 'Terravic Facial IR Database' contains various images of 20 persons. Totally there are 24508 images in the database. Images are captured in outdoor and indoor environments with different illuminations, artifacts, poses and expressions. 204 images are used from this database to be used in the methods. On the other hand, compute the error rates ground truths for these two databases are needed. Ground truths are provided by manually segmented images for the input images defined.

Pavlidis et al. [6] requires training before the pixel classification. So, for 'IRIS Thermal/Visible Face Database' 1784 background and 1390 skin images are

cropped from the entire images. 'Terravic Facial IR Database' also contains 1735 cropped background and 1240 skin images.

4.2   Comparison Metrics

The accuracy and the efficiency of the methods can be explained with different error rates. In this comparison, two different error rates as E_1 and E_2 are defined. These error rates are used to measure the accuracy and efficiency of the results pixel-by-pixel. E_1 is the classification error rate and the E_2 is the error rate that gives the mean of the FNR and FPR of the method results. If it is assumed that $R(m, n)$ is the output segmented image of the original image I, and $S(m, n)$ is the manually segmented equivalent of the output segmented image, then the E_1 is calculated as in the formula (4.1).

$$E_1 = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{MxN} \sum_{m=1}^{M} \sum_{n=1}^{N} R(m,n) \otimes S(m,n) \qquad (4.1)$$

K is the number of the image in the input folder (296 or 204). M is the number of the rows and the N is the number of the columns. $\otimes$ operator means the XOR operation. In this calculation rate, wrong classification rate is calculated, so from the XOR operation only the different pixel values on the same coordinates are found. The XOR results of these different pixels are obtained as '1'. All of these pixels are summed and divided by the size of the image. These error rates are calculated for each image individually and then the average for this error rate is found dividing the sum of the E_1 error rates to the total number of the input images.

Second error measure is calculated from the average of the False Positive Rate (FPR) and the False Negative Rate (FNR). The formula is as given in the (4.2).

$$E_2 = \frac{1}{k}\sum_{i=1}^{k}\frac{FPR}{2} + \frac{FNR}{2} \qquad (4.2)$$

TP, TN, FP and FN are explained based on the Table 4-1.

Table 4-1: Parameters Explanation for Face Segmentation

		SEGMENTED OUTPUT IMAGE	
		FACE PIXEL	BACKGROUND PIXEL
GROUND TRUTH	FACE PIXEL	True Positive	False Negative
	BACKGROUND PIXEL	False Positive	True Negative

As seen from the Table 4-1, there are two images when calculating the error rates, one is the segmented output image and the other is the manually created ground truth for the current segmented output image. The explanations are as defined below:

True Positive (TP): Defined as face pixel on segmented output image; also segmented as face pixel on ground truth,

False Negative (FN): Defined as background pixel on segmented output image; but segmented as face pixel on ground truth,

False Positive (FP): Defined as face pixel on segmented output image; but segmented as background pixel on ground truth,

True Negative (TN): Defined as background pixel on segmented output image; also segmented as background pixel on ground truth.

FPR (False Positive Rate), is the probability that background pixel in the ground truth is defined wrongly as face pixel in the segmented output image. This rate is found as in the (4.3).

$$FPR = \frac{FP}{FP + TN} \qquad (4.3)$$

FNR (False Negative Rate), is the probability that face pixel in the ground truth is defined wrongly as background pixel in the output segmented image. This rate is found as in the (4.4).

$$FNR = \frac{FN}{FN + TP} \qquad (4.4)$$

4.3 Results

This part is divided into two sections according to the database segmentation. First part includes the results with the IRIS database and the second part includes the results with the Terravic database.

4.3.1 IRIS Database Results

In the following part, IRIS database results for all methods are given in detail including both of the successful and unsuccessful results.

4.3.1.1 Pavlidis et al. [6] Method

IRIS database contains images captured in the inside. So, the temperature difference between the face and background pixel may not be sharp. So, classification of the face and background pixel may be more difficult.

Pavlidis et al. [6] used the Mixture of Gaussian model to classify the face and background pixels. This method requires training. Before obtaining the segmented results, model is trained with various skin and background models. The result for is as shown in Figure 4.1.
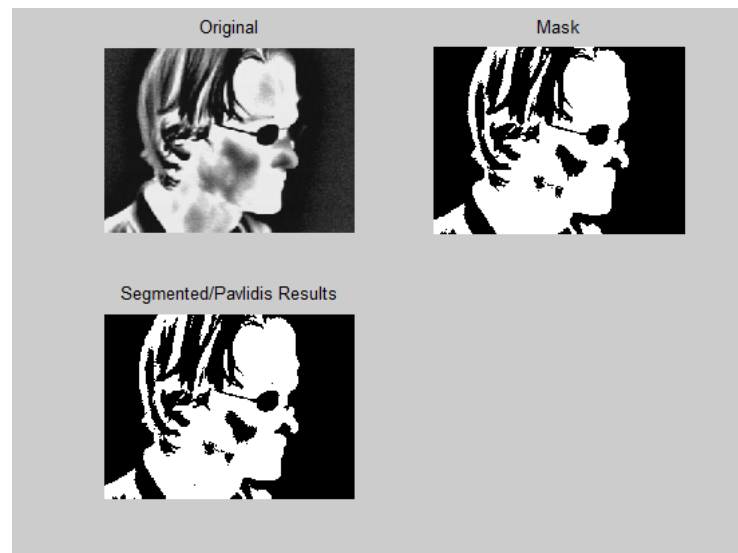

Figure 4.1 Pavlidis et al. [6] Segmentation Method Result-1/IRIS

Pavlidis et al. [6] segmentation method enables to classify the hair as the background pixel. Some part of the clothes is also segmented as background. Artifacts like glasses are not classified as face pixel. So, False Negative Rate of this method is lower. As seen from the Figure 4.1, face consists of different intensity values. Some dark regions on the face segmented as face but some parts are segmented as background. It provides improvement for some intensity values on the

face but for the regions that are so dark, segmentation process may not work so accurately. Also, clothes have similar intensity values with face. So, the classification of these cloth regions as background is difficult. So, False Positive Rate value for this method may increase because of these wrong classifications.

In the Figure 4.2, the nose is very cold and its intensity value is very low when compared with the other parts of the face. In this circumstance, it is not possible to segment nose as a part of the face in this method because its intensity is not in the defined distribution range for the face. But some of these foreground pixels which have lower intensity values are classified properly as face if their intensity values fall into the bimodal Gaussian model.

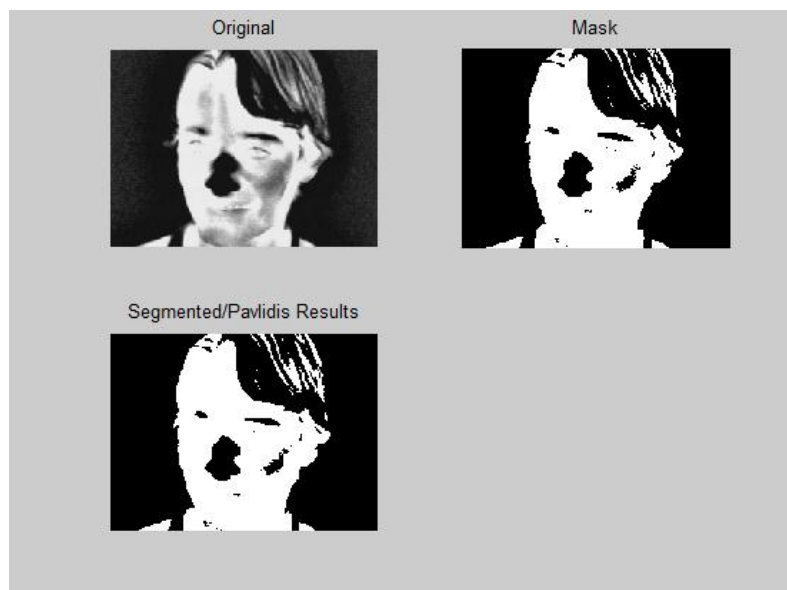Error rates for the results of this method for IRIS database are; E_1 is 0.176 and E_2 is 0.169.

Figure 4.2 Pavlidis et al. [6] Segmentation Method Result-2/IRIS

4.3.1.2 Cho et al. [19] Method

This method based on the edge detection and morphological operations. This method is fast when compared with other methods. The segmented results are given as in the Figure 4.3 and Figure 4.4 titled as difference image. In Figure 4.3, the face segmented successfully from the image. Hair and clothes are labeled as background. On the other hand, as mentioned before, the main purpose of the face segmentation is the elimination of the shoulders and the neck from the image. From these results, neck is included in the segmented images. So, this method is not enough to eliminate the neck from the segmented part of the image.
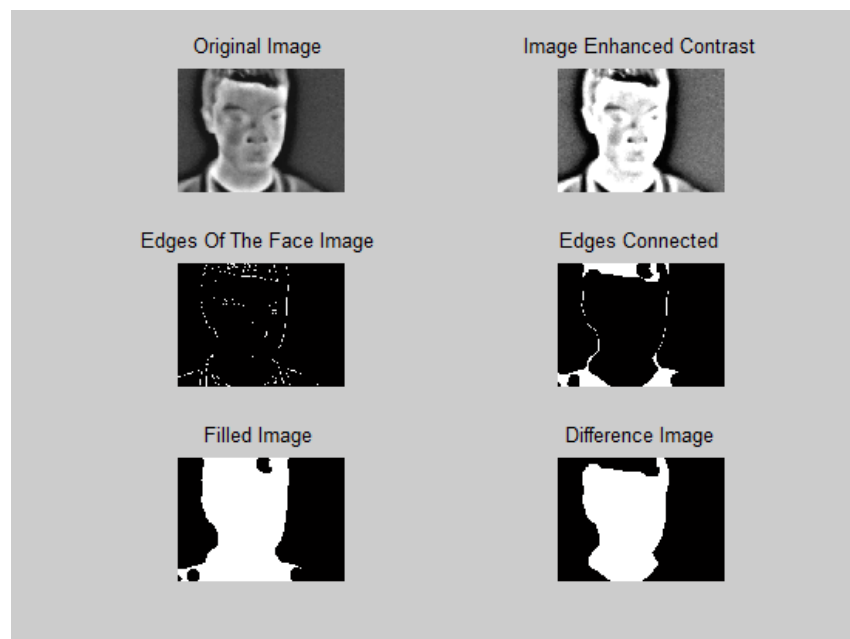

Figure 4.3 Cho et al. [19] Segmentation Method Result-1/IRIS

In Figure 4.4, hair, clothes and the glasses are eliminated from the image successfully; but some background part of the image is segmented as face. The result of this wrong classification caused the high increases in the FPR values. The reason for this wrong segmentation is the unclosed contours. When the biggest

contour which is chosen as face contour is not closed, some background part of the image is also segmented as face after filling the contour.

Error rates for the results of this method for IRIS database are; E_1 is 0.309 and E_2 is 0.296.
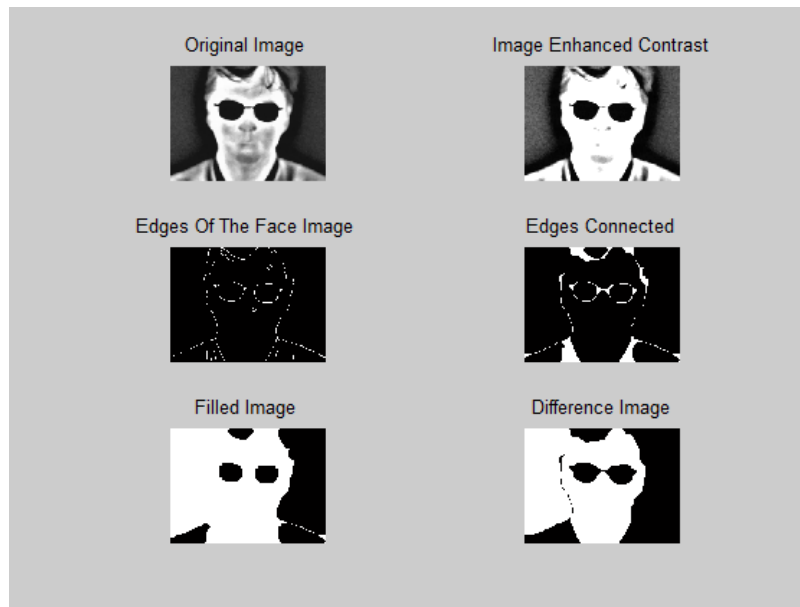

Figure 4.4 Cho et al. [19] Segmentation Method Result-2/IRIS

4.3.1.3 Filipe et al. Method [24]

This method uses different approaches together to overcome the deficiencies of other methods. Elimination of the clothes and the neck is the one of the most important feature of this method. The results are shown in Figure 4.5 and Figure 4.6.
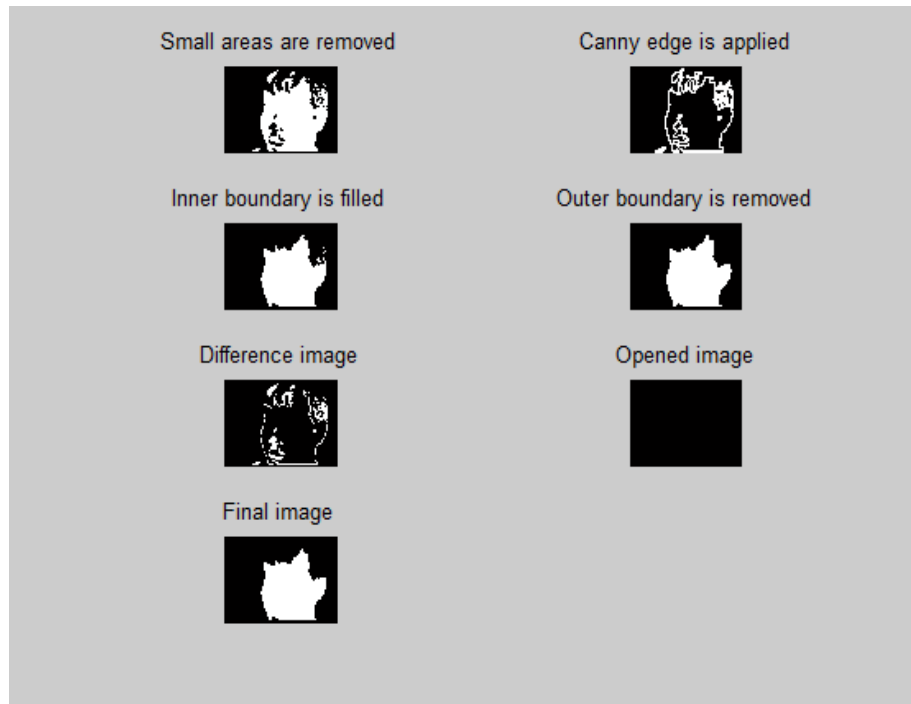
Figure 4.5 Filipe et al. [24] Segmentation Method Result-1/IRIS

Figure 4.5 shows the steps after active contour method. Final image is the segmented face image. The main motivation of this method is to exclude the clothes, shoulder and the neck from the segmented image. So, as seen from the Figure 4.5, hair, clothes, and the neck parts are not included in the segmented face. So, FPR value for this method is lower. The results only include the face part of the image.

Figure 4.6 shows a contribution of this method. After active contour method, the image is obtained as in the first picture of the Figure 4.6. Then, the biggest contour is full-filled as seen in the Figure 4.6 which has the title 'Inner boundary is filled'; but glasses are classified as face pixels. So, to correct this wrong classification, difference between the full-filled image and the image segmented after active contour method is found. Basically, glasses are included in the difference image with other redundancy parts. So redundancies are eliminated and only the glasses obtained from the difference image. Then OR operation is applied to the glasses and the full-filled image. Therefore, the segmented face image took its final form as seen

from the Figure 4.6. On the other hand, glasses and a circular pixel group remained in the image. To avoid this, opening function may be operated with a bigger morphological operator; but the change of the morphological operator may affect other segmentation process negatively.

Active contour is an efficient way to find the edges but if the initial point is not well defined, active contour may fail while converging to the edges. Because of this reason, the selection and the determination of the initialization point is very important. And also, down line of the face limit may fail while calculating the best-fitting curve. Best-fitting curve may not be fitted to the chin or the shoulders always. In this case, down line of the face limit will not give the expected results and this will affect the following segmentation part of the method.
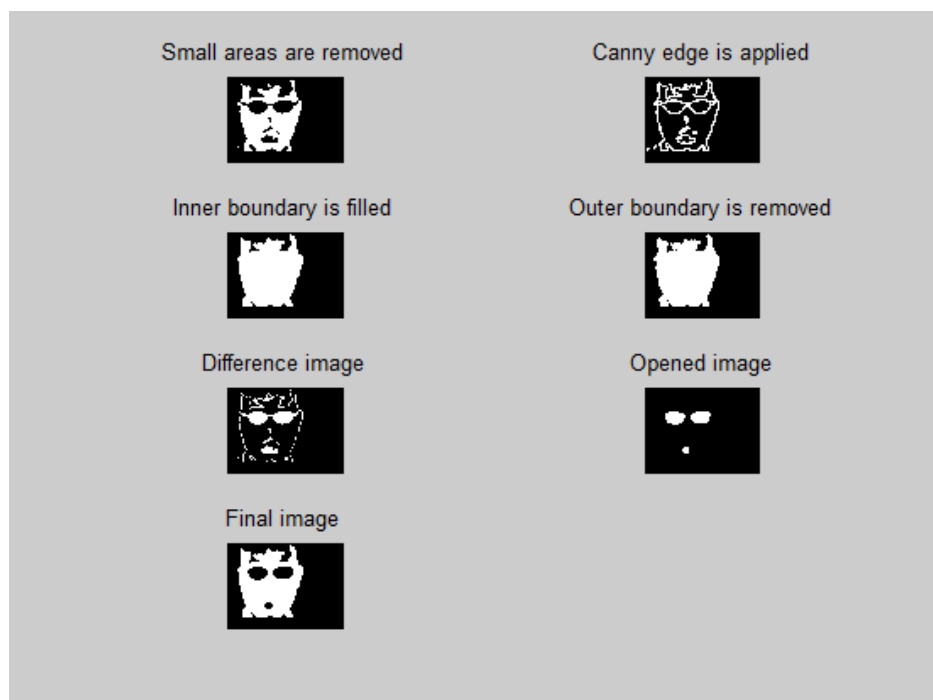

Figure 4.6 Filipe et al. [24] Segmentation Method Result-2/IRIS

It is an important point that, Filipe et al. [24] used 235 images from the Terravic database; but in the implementation, only 204 images in this dataset are used because of the limited downloading access to these databases.

On the other hand, the wrong determination of the face limits can cause inefficient segmentation results. Face limits are found by the sharp changes in the intensity levels. If there is a sharp intensity change on the face, then the limits will be defined at these coordinates. Thus, face limits may be determined wrong. This may also effect the segmentation, because the coordinates of the initialization ellipse for the active contour is going to change at first.

Error rates for the results of this method for IRIS database are; E_1 is 0.111 and E_2 is 0.139.

4.3.1.4   Filipe et al. Method [32]

This method is based on the Pavlidis et al. [6] method. The main deficiency of Pavlidis et al. [6] method is the inclusion of the background pixels like clothes in the segmented part of the face. To overcome this deficiency a region of interest is defined from the boundaries of the face using signature values. But it is important to mention that this method is only applied to the FSU and UND databases in the original paper [5]. On the other hand, in this implementation, this method is applied to the IRIS and Terravic databases.

This method is used the signature approach to eliminate the clothes and the hair from the image. In Figure 4.7 the face is totally included in the segmented part. It is observed that FNR value of this method for the IRIS database is lower than the FNR value of the Pavlidis method.
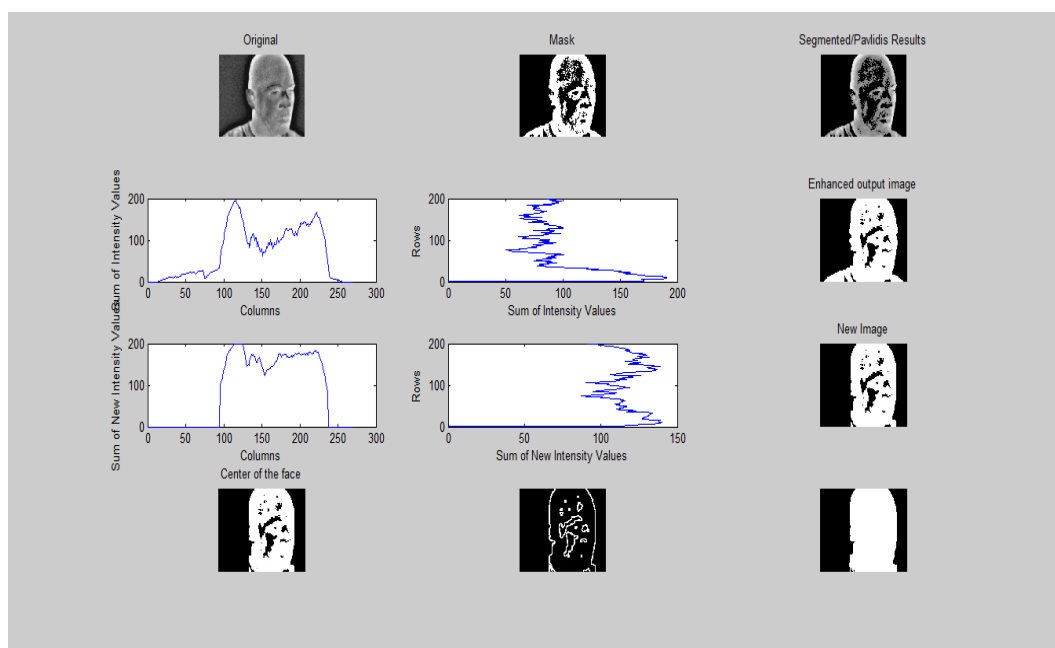
Figure 4.7 Filipe et al. [32] Segmentation Method Result-1/IRIS

Signature approach provides to obtain only the face part as much as possible; but as seen in Figure 4.8, signature of the row which includes the glasses are equalized to zero, because the signature value of this row exceeded the defined threshold value. So, the segmentation gave an invalid output as seen in Figure 4.8.

As it is seen from the results, signature approach may be inefficient when a threshold value is defined; because this approach is sensitive to the glasses or moustache.

Another deficiency of this method is the existence of the unclosed contour. The contour which is extracted may be unclosed and as a result of this, segmented part may include much of the background as a part of the face also. So, FPR value is increased because of the unclosed contours in this method when compared with the Pavlidis method FPR values.

62

Error rates for the results of this method for IRIS database are lower than the Pavlidis mathod; E_1 is 0.172 and E_2 is 0.143.
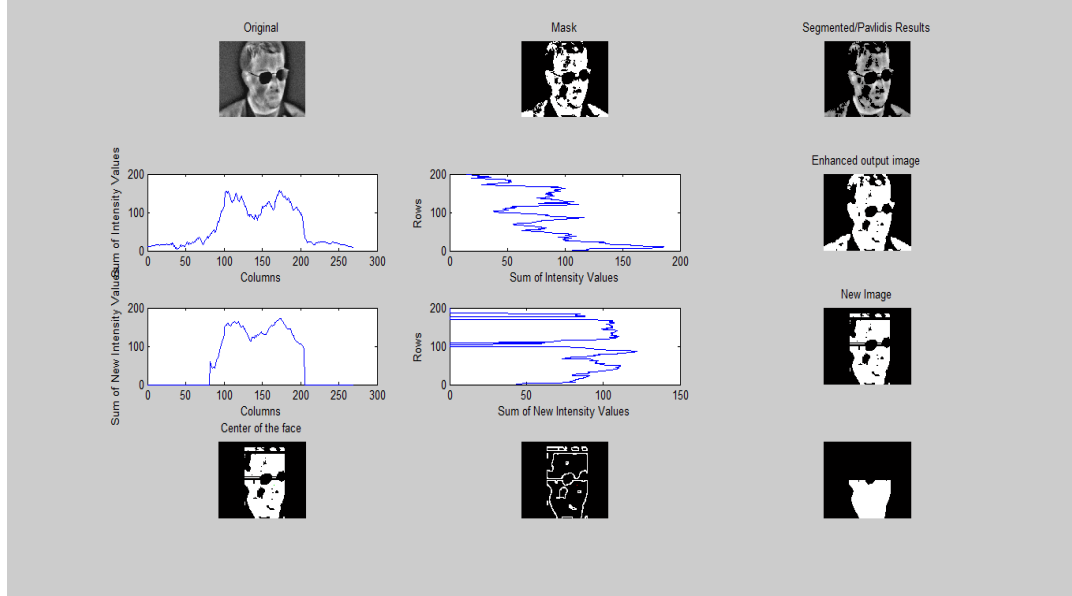

Figure 4.8 Filipe et al. [32] Segmentation Method Result-1/IRIS

4.3.2 Terravic Database Results

Segmentation results in this database are lower for all methods. Terravic database contains various images captured in the indoor and in the outdoor. Outdoor images have higher contrast values. Therefore, segmentation in this database gives more accurate results. In the following part, Terravic database results for all methods are given in detail including both of the successful and unsuccessful results.

4.3.2.1 Pavlidis et al. [6] Method

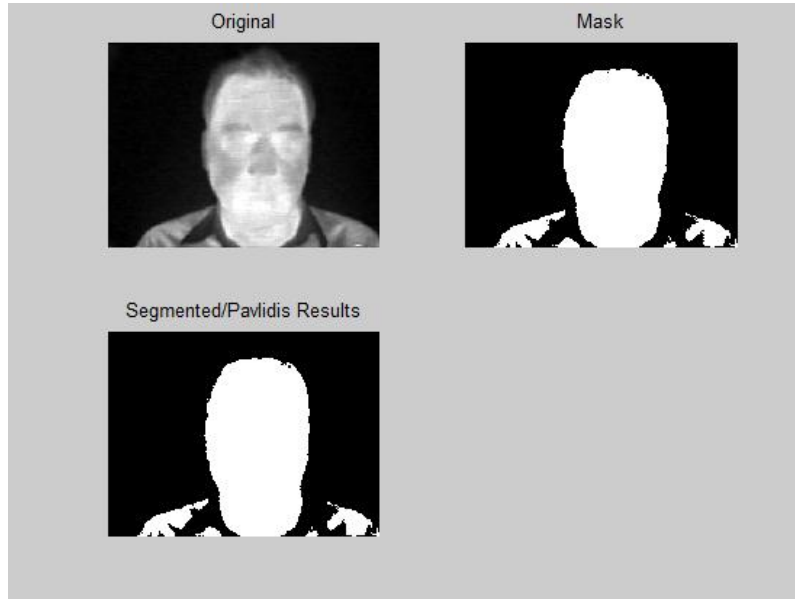Results for the Terravic database are given in the Figure 4.9 and Figure 4.10.

Figure 4.9 Pavlidis et al. [6] Segmentation Method Result-1/Terravic

Figure 4.9 shows the segmentation result for Terravic database. As seen from the Figure 4.9, hair is not classified as face and also some part of the clothes is classifed as background. However, neck is also included in the segmented face. But totally, face part is segmented more accurately for this database with this method than the IRIS database.

Figure 4.10 is an example to the hair and artifact (glasses) exclusion. This method is pose-invariant.

Error rates for the results of this method for Terravic database are; E_1 is 0.099 and E_2 is 0.075.
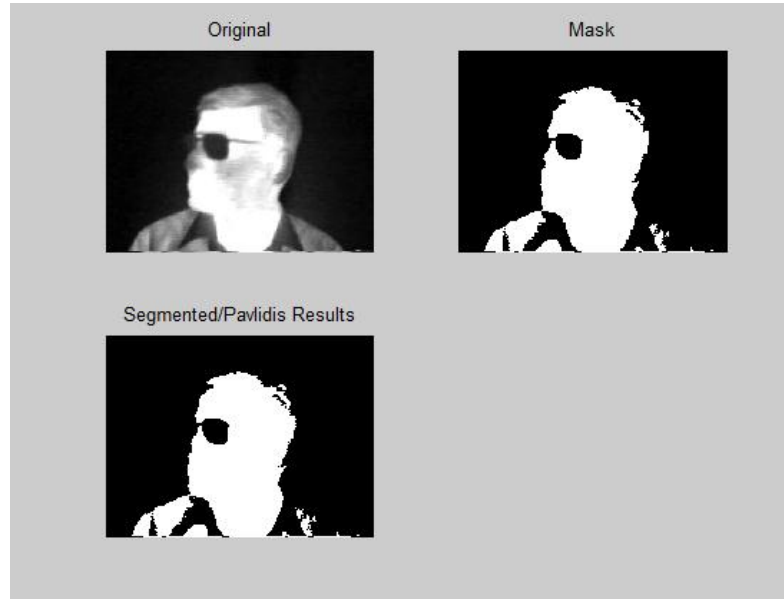
Figure 4.10 Pavlidis et al. [6] Segmentation Method Result-2/Terravic

4.3.2.2   Cho et al. [19] Method

The results for this method with Terravic database are given in Figure 4.11, Figure 4.12, Figure 4.13 and Figure 4.14. In Figure 4.11, segmentation result is accurate, shoulder, clothes and hair are discarded; but neck is included. This method may fail if the contour is not closed. An example of this is shown in Figure 4.12. Background part is also classified as face. In Figure 4.13, face is discriminated from the artifact (glasses). Shoulders and the clothes are also excluded; but neck is included. Figure 4.14 is an example for the wrong segmentation. The reason for this is the wrong contour formation.

Error rates for the results of this method for Terravic database are; E_1 is 0.215 and E_2 is 0.159.
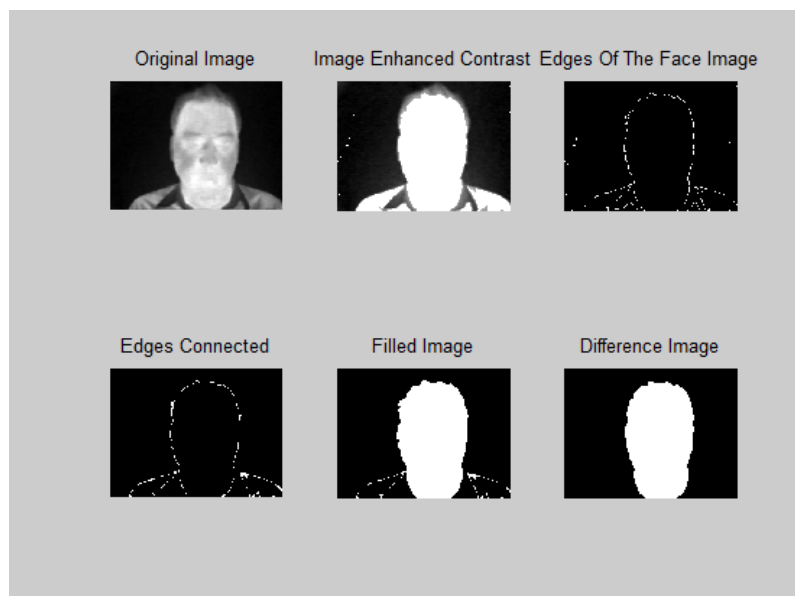
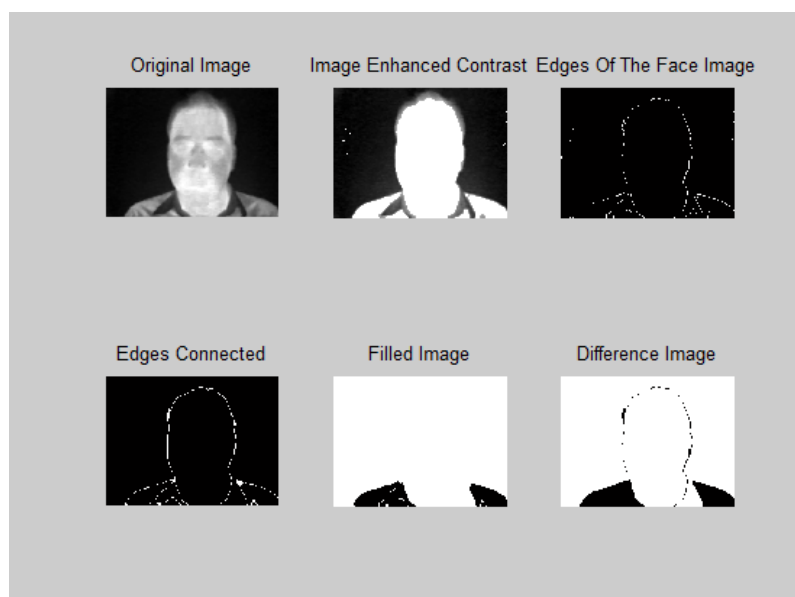Figure 4.11 Cho et al. [19] Segmentation Method Result-1/Terravic

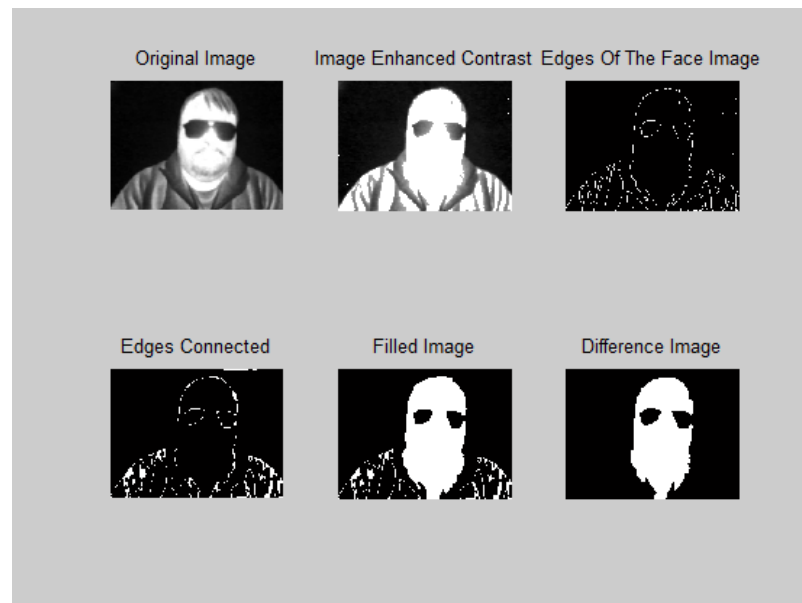Figure 4.12 Cho et al. [19] Segmentation Method Result-2/Terravic

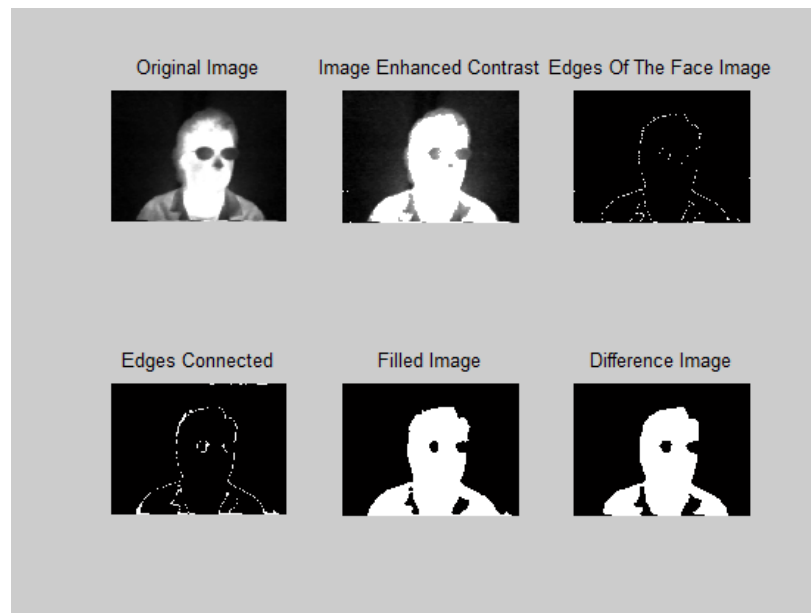Figure 4.13 Cho et al. [19] Segmentation Method Result-3/Terravic

Figure 4.14 Cho et al. [19] Segmentation Method Result-4/Terravic

4.3.2.3 Filipe et al. Method [24]

This method results are shown in Figure 4.15, Figure 4.16, Figure 4.17, Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21, Figure 4.22, Figure 4.23 and Figure 4.24 . In Figure 4.15, an accurate segmentation result is given. Best-fitting curve is fitted to the shoulders and the shoulders are eliminated with the aid of this curve. Then, active contour method is applied and the face contour is defined as in the Figure 4.15.

Figure 4.15 Filipe et al. [24] Segmentation Method Result-1/Terravic

In Figure 4.16, the segmented image with active contour in Figure 4.15 is enhanced with some morphological and image enhancement technics. But for some cases, chin part of the person may be eliminated when the curve is fitted to the shoulders. So, this will result to the increase in the FNR value of the method.

Figure 4.17 is an example for the wrong segmentation. A seen from the figure, best-fitting curve is not found in the image. So, shoulders and the neck are included in the segmentation. And also, when there is no limitation for the active contour method,

the method may converge to irrelevant parts on the image or may stick into hair, artifacts or background. So, the segmentation result will not be efficient as seen in the Figure 4.18. Absence of the best fitting curve causes the increase on the FPR.

In Figure 4.19, shoulders are not on the same level because of the position of the person. But, best-fitting curve still works in this condition. Hat artifact is also eliminated and the final form of the segmented image is as shown in Figure 4.20.

Figure 4.21 shows the best-fitting curve fitted to the chin. So, the remaining part below the image from the chin is assumed as background. The segmentation result for the Figure 4.21 is shown in Figure 4.22.

Another example for chin-fitting curve is given in Figure 4.23. In this figure, segmentation is done with the existence of the artifact like glasses. The segmentation result is given in the Figure 4.24.

Error rates for the results of this method for Terravic database are; E_1 is 0.066 and E_2 is 0.068.
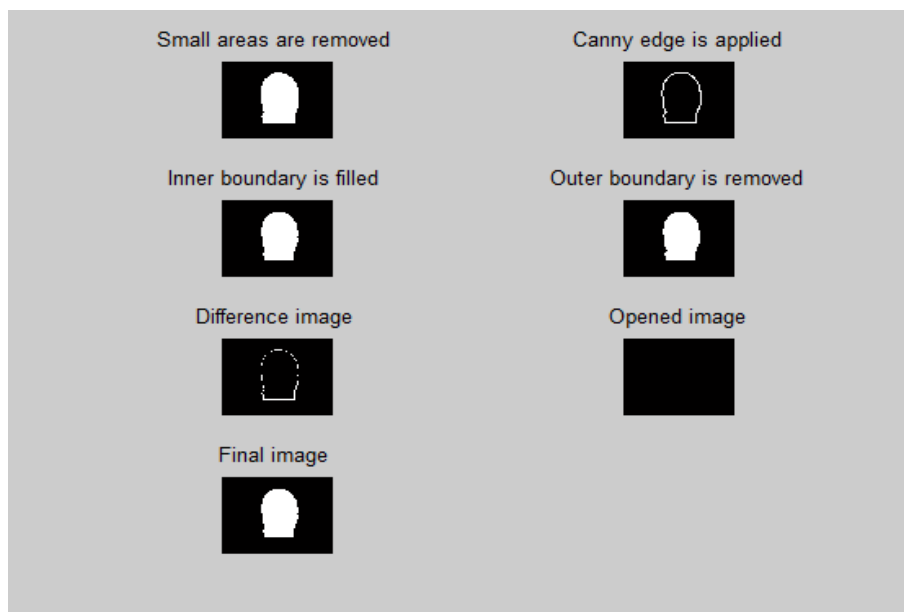
Figure 4.16 Filipe et al. [24] Segmentation Method Result-2/Terravic

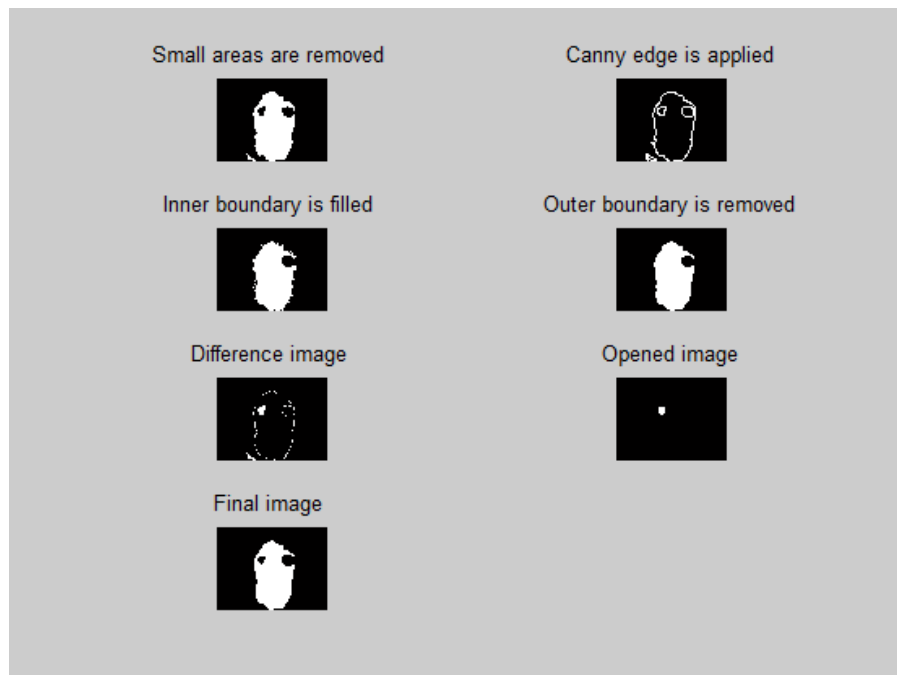Figure 4.17 Filipe et al. [24] Segmentation Method Result-3/Terravic

Figure 4.18 Filipe et al. [24] Segmentation Method Result-4/Terravic

Figure 4.19 Filipe et al. [24] Segmentation Method Result-9/Terravic

Figure 4.20 Filipe et al. [24] Segmentation Method Result-10/Terravic

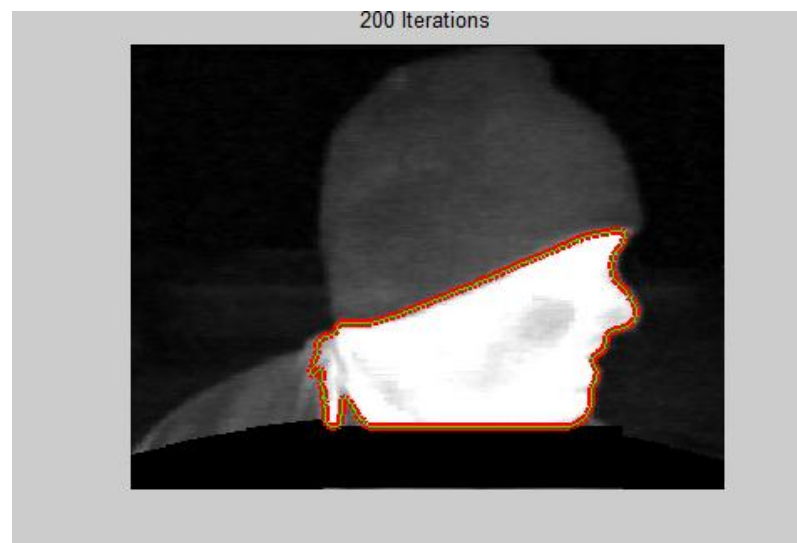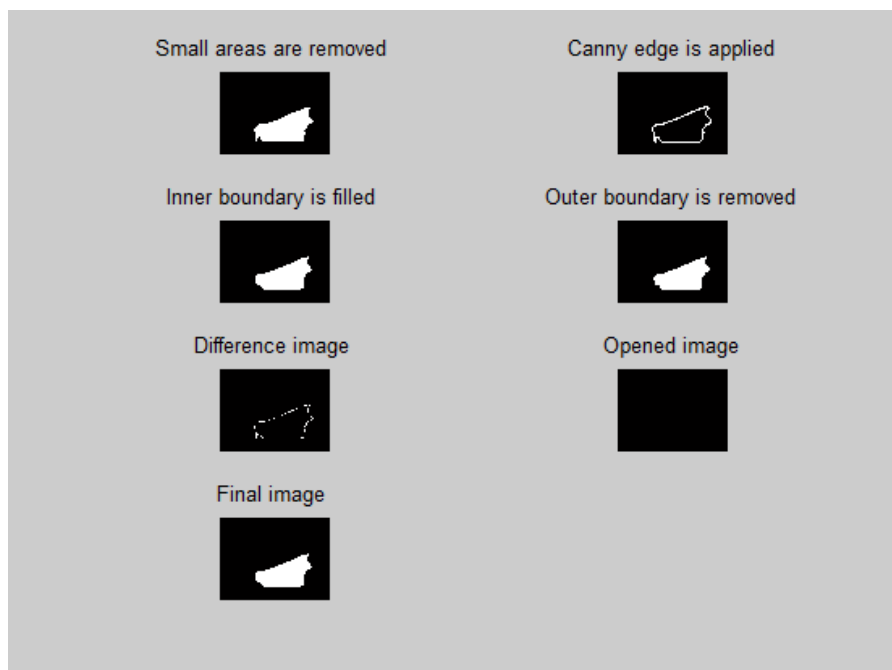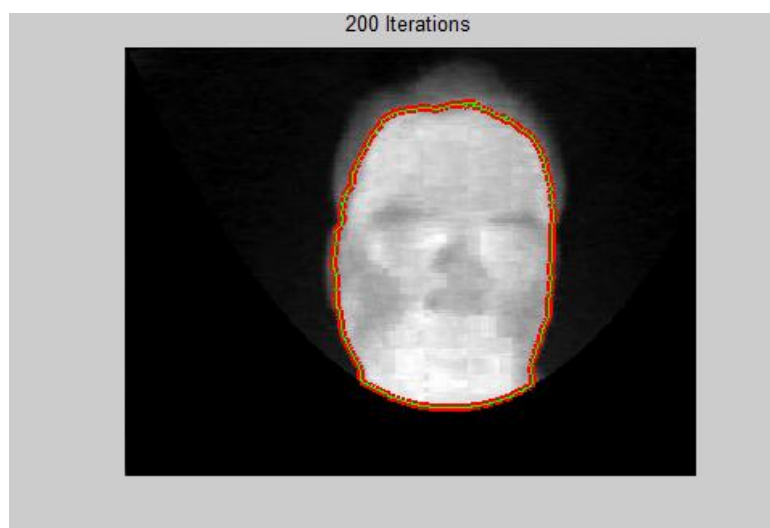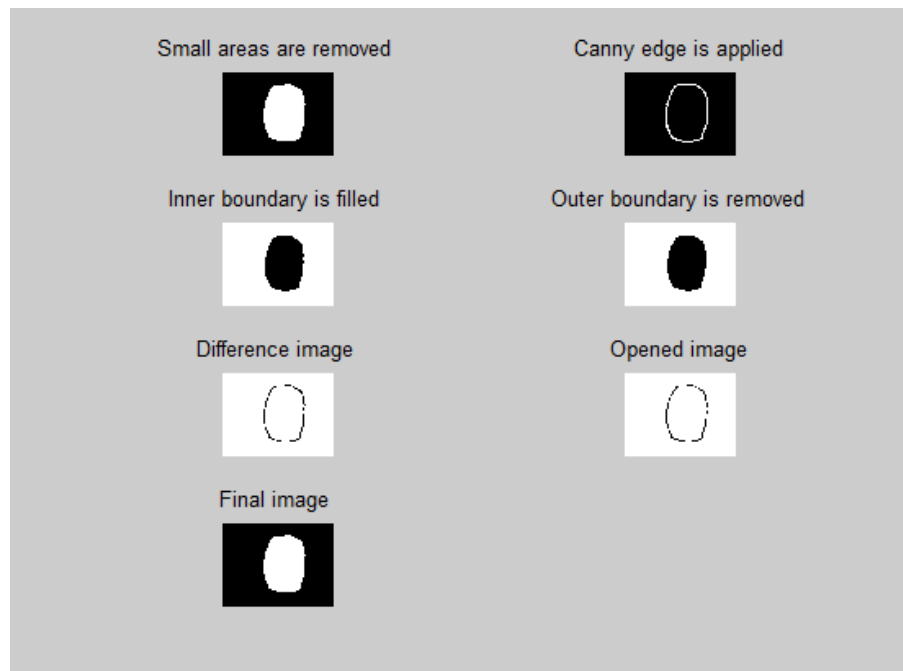Figure 4.21 Filipe et al. [24] Segmentation Method Result-11/Terravic

72

Figure 4.22 Filipe et al. [24] Segmentation Method Result-12/Terravic

Figure 4.23 Filipe et al. [24] Segmentation Method Result-13/Terravic
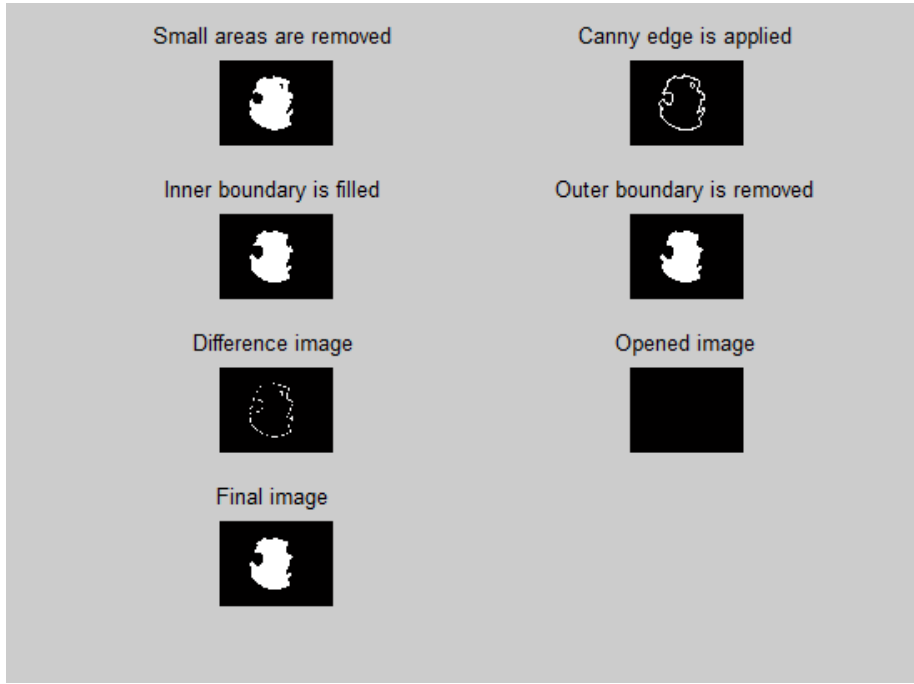
Figure 4.24 Filipe et al. [24] Segmentation Method Result-14/Terravic

4.3.2.4   Filipe et al. Method [32]

As seen from the Figure 4.25, this method is sensitive to the artifacts like glasses. The image is divided from the location of the glasses, because the intensity value of this row is higher than the defined threshold value. So, the face is segmented partially. This causes the increase in the FNR rate when compared with the Pavlidis results.

In Figure 4.26, segmentation is achieved successfully. The clothes are excluded; but the neck is again included in the segmented face part.

Error rates of this method for Terravic database are so close to the Pavlidis error rates. E_1 error rate is smaller but E_2 is very close to the Pavlidis error rates. Error rates are; E_1 is 0.097 and E_2 is 0.078. The reason is the less contribution of the

clothes in Terravic database. When the clothes are not classified as face pixels in Pavlidis method, this method may not provide so much improvement.
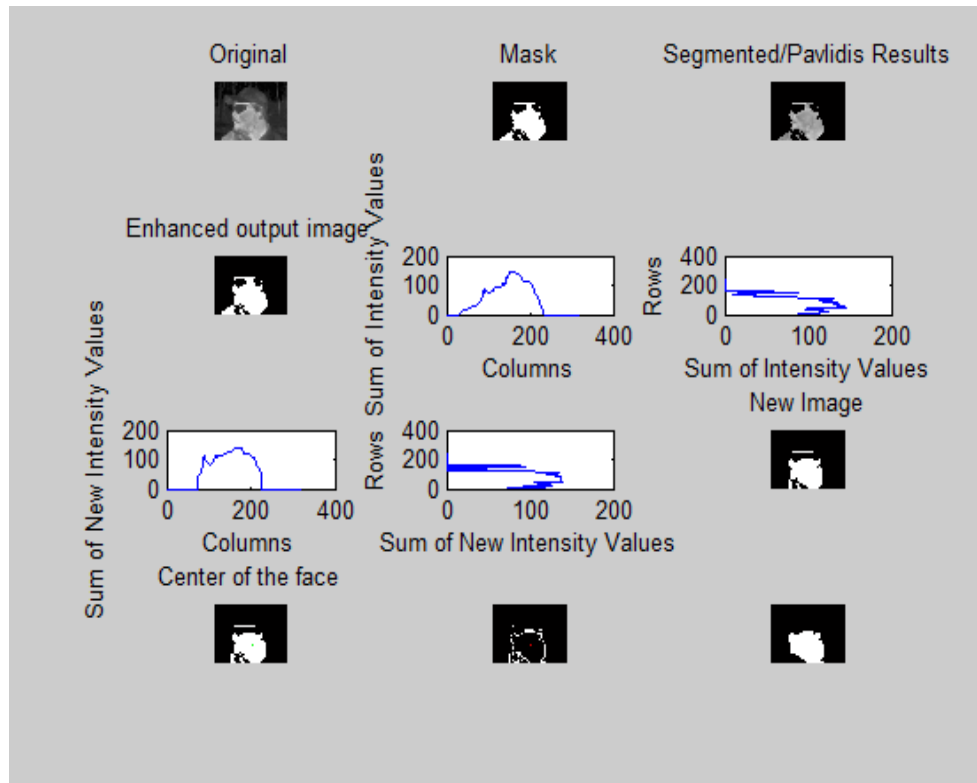
Figure 4.25 Filipe et al. [32] Segmentation Method Result-1/Terravic
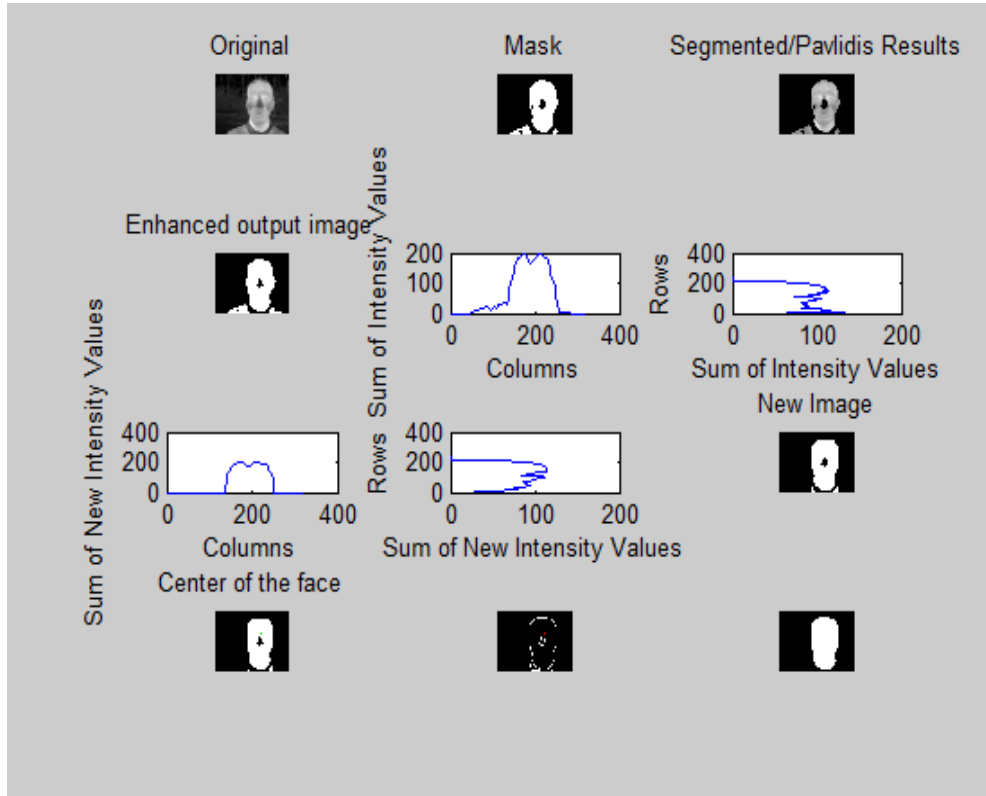
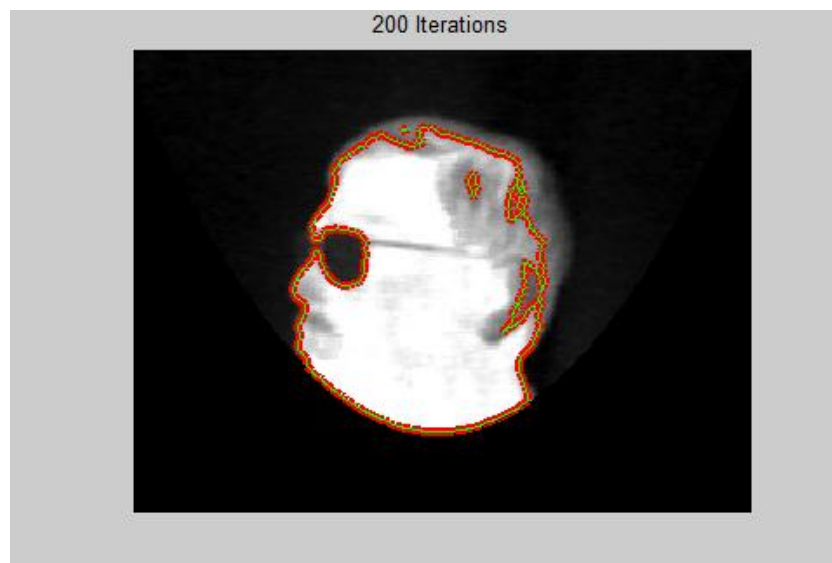Figure 4.26 Filipe et al. [32] Segmentation Method Result-2/Terravic

## 4.4 Comparison Of The Results and The Discussion

In this part of the thesis, computational and visual results obtained from IRIS and Terravic databases are explained respectively.

4.4.1 Computational Comparison

Comparison table for all methods in all databases are given in Table 4-2. Each method is compared for two databases; IRIS and Terravic. Each parameter is given for one image.

Table 4-2 Comparison of the methods

		METHODS			
		Cho et al. [19]	*Pavlidis et al.* [6]	*Filipe et al.* [24]	*Filipe et al.* [32]
IRIS Database	E_1	0.309	0.176	**0.111**	0.172
	E_2	0.296	0.169	**0.139**	0.143
	FNR	0.262	0.157	0.200	0.081
	FPR	0.332	0.182	0.078	0.202
	Time (s)	0.9640	4.79	9,64	5.24
Terravic Database	E_1	0.215	0.099	**0.066**	0.097
	E_2	0.159	0.075	**0.068**	0.078
	FNR	0.074	0.038	0.070	0.05
	FPR	0.245	0.112	0.065	0.107
	Time (s)	0.7303	3.301	6.64	3.68

As seen from the Table 4-2, Filipe et al. [24] provides the best error rates for all databases. Cho et al. [19] has the higher error rates among other methods. The reason for this is the unclosed contours in the image. The unclosed contours cause to increase in the False Positive Rate. Pavlidis et al. [6] has lower error rates than Cho et al. [19] and higher error rates than Filipe et al. [24]. Pavlidis method is based on the distribution model of the pixel intensity values. So, all warm parts of the clothes

and all cold parts of the face could not be segmented correctly and it caused wrong segmentation results. However, Filipe et al. [32] is developed to improve the segmentation efficiency of Pavlidis et al. [6] method. Error rates are improved except the Filipe et al. [32] Terravic E_2 error rate. Majority of the cloth pixels are classified as background in Terravic database using the Pavlidis method, therefore the improvement of Filipe et al. [32] method is not apparent in Terravic database when compared with the IRIS database error rates. So, Filipe et al. [24] solved this problem using best-fitting curve for the chin or for the shoulders. Also, the active contour provided the accurate segmentation for the remaining part of the image. But as a short-coming of this algorithm, best fitting curve is not found for some images or active contour method is stacked into the detail.

Terravic database has lower error rates for each method. The reason is the high contrast of the images in this database. This database is captured in indoor and outdoor. So, segmentation is not difficult as it is in the IRIS database. Also, Pavlidis et al. [6] method provides lower error rates for this database. The reason is the less contribution of the clothes in the image. Contrast value is higher than the IRIS database. So, the segmentation is much more accurate than the IRIS database. But it is observed that the method requires much more execution time for IRIS database. The reason for this is the total processed pixels in IRIS database are much more than the total processed pixels in Terravic Database. So execution time for IRIS database takes longer time.

By downsampling the images, the execution time may be shorten. It is observed that downsampling from 320x240 to 160x120 in IRIS database is shortened the execution time approximately by %20.

Cho et al. [19] and Filipe et al. [32] provide the shoulder elimination for some images but neck is always included in the images. So, this causes to have higher FPR values in IRIS database.

The method used in Filipe et al. [24] may be deceptive for some conditions. For example when the skin intensity is very high, it is difficult to determine the curve fitting to the chin. It may be possible to detect curve fitting to the shoulders; but if both of the curves could not be determined, the segmentation result may not be sufficient.

4.4.2   Visual Comparison

In this part, obtained outputs for all of these four methods are compared visually.
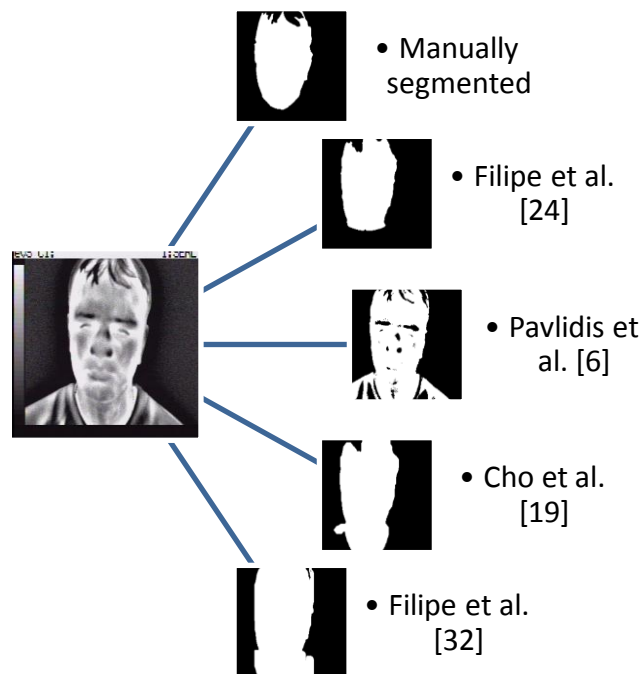
Figure 4.27 Comparison of methods visually-1/IRIS Database

79

• Manually segmented

• Filipe et al. [24]

• Pavlidis et al. [6]

• Cho et al. [19]

• Filipe et al. [32]
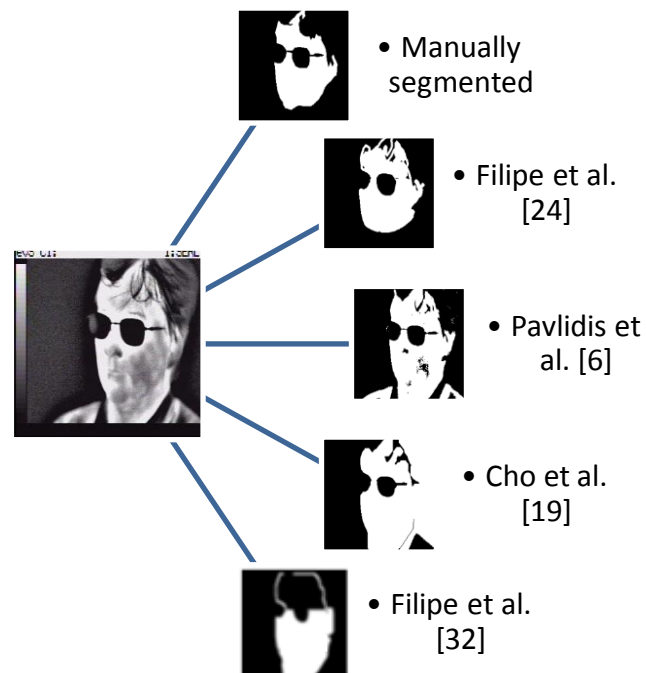
Figure 4.28 Comparison of methods visually-2/IRIS Database

• Manually segmented

• Filipe et al. [24]

• Pavlidis et al. [6]

• Cho et al. [19]

• Filipe et al. [32]
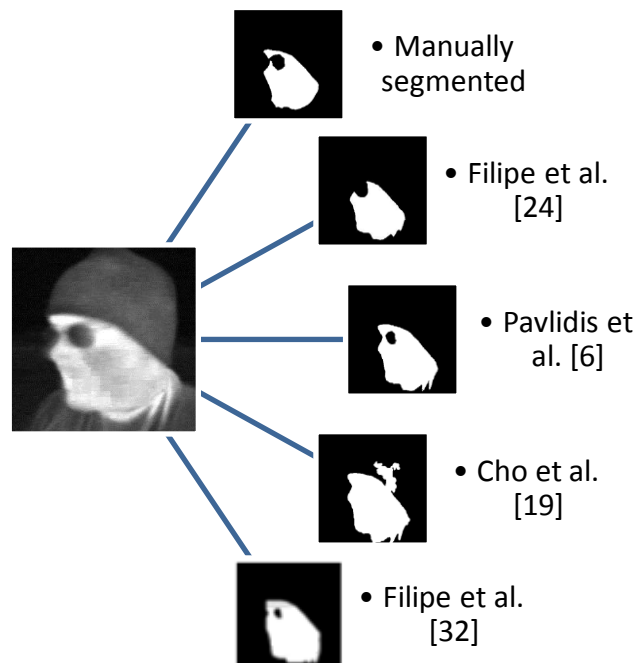
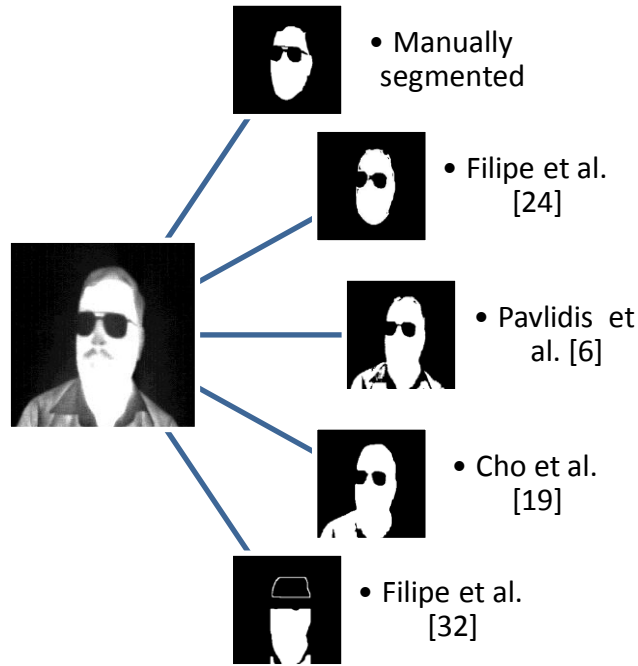Figure 4.29: Comparison of methods visually-1/Terravic Database

Figure 4.30 Comparison of methods visually-2/Terravic Database

As seen from Figure 4.27, Figure 4.28, Figure 4.29 and Figure 4.30, Filipe et al. [24] provides the best results for the face segmentation. It provides the nearest results to the manually segmented images. Some part of the hair may be included in the segmented face part but it is a very small part. Pavlidis et al. [6] includes clothes as foreground pixels in the segmented part. Also, this method labeled some part of the face as the background. Cho et al. [19] includes the neck in the segmented part. Filipe et al. [32] method provides the elimination of the clothes in Figure 4.27; but hair and the neck is included in the segmented part.

In Figure 4.28 Cho et al. [19] includes the background as segmented part because of the unclosed contour. As mentioned before, unclosed contours may cause wrong classifications of background pixels as face pixels. On the other hand, in Filipe et al. method [32], glasses caused the row where the glasses are located to be equalized to zero. As a result of this, face is divided into two regions and only the part which includes the center of the face is segmented as face.

In Figure 4.29, the most similar result to the manually segmented face is provided by the Filipe et al. [24]. Cho et al. [19] method is segmented some part of the background (some part of the hat) as face pixels. Some part of the glasses is classified as face pixels in Filipe et al. [32] method because of the morphological operations.

It is mentioned that if the skin contrast is similar in the whole skin part, it is very difficult to differentiate the chin or shoulder from the image. However, in Figure 4.30, Filipe et al. [24] provided an accurate result with the exception of the neck using curve fitting method. On the other hand, moustache and cold nose are labled as face pixels in Pavlidis et al. [6] method; but clothes are considered as face pixels as seen in the Figure 4.30. In Cho et al. [19] method, shoulders are not classified accurately because of the unclosed contours. The sensitivity to the glasses is also negatively affected the face segmentation results in Filipe et al. [32] method.

In conclusion, it is observed that Filipe et al. [24] provides the best results for face segmentation among other implemented methods. It provides curve fitting to overcome the shoulder or neck inclusion on the segmented part of the image. Also, provides active contour method to segment image more accurately. Signature values are also used in the method to improve the effectiveness of the active contour method; because the location of the initialization point is critical to obtain accurate face segmentation results.

As seen from the error rates comparison or the visual comparison, Filipe et al. [24] is an efficient face segmentation method to be implemented in face recognition methods.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

In this chapter conclusions and the future works are explained. The works can be done are described on the future works.

5.1   Conclusions

In this study, different methods for IR face segmentation have been analyzed. The aim is to compare the results in terms of computation and accuracy.

The results are compared on IRIS and Terravic databases. It is observed that the Filipe et al. [24] provides the lowest error rates both for the E_1 and the E_2 in both databases.

The reason for this is the robustness of the method. Filipe et al. [24] method is pose-invariant. The method is able to exclude the clothes or other artifacts like glasses, hats etc. from the segmented part of the image. Other methods like Cho et al. [19] or Filipe et al. [32] also focused on the elimination of the clothes from the image but their results are not satisfactory. One of the important features of the Filipe et al. [24] method is its efficiency in finding the face contour. In this method, face contours are found dynamically using active contour method. In other methods, face contour are found using classical methods like Canny or Sobel edge detection technics. Also, one of the most important features of the Filipe et al. [24] method is the LSE algorithm which provides the elimination of the shoulders and the neck

from the image. LSE finds the best chin or shoulder curve to be fitted. On the other hand, signature approach is also substantial for the Filipe et al. [24] method, because signatures forms the bounding box for the face and determines the initialization point for the active contour method. In other implemented methods, a threshold value is defined for the signatures to limit the face regions; but it caused some problems when the glasses are exist in the image. On the other hand, extrema points of the signatures are found from the first derivative of the function by Filipe et al. [24]. Therefore, considering all of these features and comparing the computational and visual results, it is analyzed that the Filipe et al. [24] is an optimal and functional segmentation method for recognition processes.

5.2 Future Works

In this study IRIS and Terravic databases have been used. These databases include only one face within the image; but images including multiple faces may be used for the multiple face segmentation. It is important to be able to segment multiple faces in real world applications. Therefore, one of the improvements which can be suggested is to be able to segment multiple faces within an image. In addition, methods can be extended to segment faces from the videos by incorporating different algorithms in the segmentation stage.

REFERENCES

[1] **Ross, A., Nandakumar, K., Jain, A.** *Handbook of multibiometrics.* New York : Springer, 2006.

[2] **Jain A., Flynn P., Ross A.** *Handbook of Biometrics.* New York : Springer, 2007.

[3] **Filipe, S., Alexandre, L.A.** *Algorithms for invariant long-wave infrared face segmentation: evaluation and comparison.* s.l. : Springer, 2012. pp. 823-837.

[4] **Moreno, M. M.,.** *Reconocimiento Biométrico Basado En Imagénes Simuladas En La Banda de Ondas Milimétricas.* Madrid : http://arantxa.ii.uam.es/~jms/pfcsteleco/lecturas/20120614MiriamMorenoMoren o.pdf, 2012.

[5] **Akhloufi, M. A.,.** *Reconaissance Des Visages Par Imagerie Multispectral.* 2013.

[6] **Pavlidis, I., Buddharaju, P., and Manohar, C., Tsiamyrtzis, P.,.** Biometrics: Face Recognition in Thermal Infrared. *Biomedical Engineering Handbook.* s.l. : CRC Press, 2006, Vol. 3, pp. 1-15.

[7] [Online] 04 10, 2014. http://socia-lab.di.ubi.pt/~silvio/IbPRIA2013.html.

[8] **Wang, H., and Chang, S.F.,.** *A Highly Efficient System for Automatic Face Region Detection in MPEG Video.* s.l. : Circuits and Systems for Video Technology, IEEE Transactions, 1997. Vol. 7, 4.

[9] **Hsu, R., Abdel-Mottaleb, M. and Jain, A.,.** *Face detection in color images.* s.l. : IEEE Trans. Pattern Analysis and Machine Intelligence, 2002. Vol. 24, 5, pp. 696-706.

[10] **Sobottka, K., Pitas, I.** *Segmentation and tracking of faces in color images.* s.l. : IEEE, 1996. pp. 236-241.

[11] **Lin, Hwei-J., Yen, Shwu-H., Yeh, Jih-P., and Lin, Meng-J.** *Face Detection Based on Skin Color Segmentation and SVM Classification.* Yokohama : IEEE, Secure System Integration and Reliability Improvement, 2008. SSIRI '08. Second International Conference on, 2008. pp. 230 - 231.

[12] **A. Alabbasi, H., Moldoveanu, F.** *Human face detection from images, based on skin color.* Sinaia : IEEE, System Theory, Control and Computing (ICSTCC), 18th International Conference, 2014. pp. 532 - 537.

[13] **Kwolek, B.,.** *Face Tracking System Based on Color, Stereovision and Elliptical Shape Features.* s.l. : IEEE Inter Conf on Advanced Video and Signal Based Surveillance, 2003. pp. 21-26.

[14] **Jalilian, B., Chalechale, A.** *Face and Hand Shape Segmentation Using Statistical Skin Detection for Sign Language Recognition.* s.l. : http://www.hrpub.org, 2013.

[15] **M. Hasan, M., K. Mishra, P.** *Superior Skin Color Model using Multiple of Gaussian Mixture Model.* s.l. : British Journal of Science, 2012. Vol. 6.

[16] **Marius, D., Pennathur, S., and Rose, K.,.** *Face Detection Using Color Thresholding, and Eigenimage Template Matching.* [Online] [Cited: 01 17, 2015.] https://web.stanford.edu/class/ee368/Project_03/Project/reports/ee368group15.pdf.

[17] **Gasparini, F., Schettini, R.** *Skin segmentation using multiple thresholding.* s.l. : SPIE Proceedings, 2006.

[18] **Kumar O S, S., G Kamal, A.,.** *A Modified Algorithm For Thresholding and Detection of Facial Information From Color Images Using Color Centroid Segmentation and Contourlet Trasform.* s.l. : Signal & Image Processing : An International Journal (SIPIJ), 2011.

[19] **Cho, S., Wang, L., and Ong, W.,.** *Thermal imprint feature analysis for face recognition.* s.l. : IEEE International Symposium on Industrial Electronics, 2009. pp. 1875-1880.

[20] **Gyaourova, A., Bebis, G., Pavlidis, I.,.** *Fusion of Infrared and Visible Images for Face Recognition.,.* s.l. : 6th Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA, Springer, 2013. pp. 632-639.

[21] **Selinger, A., Socolinsky, D.,.** *Appearance-based facial recognition using visible and thermal imagery.* s.l. : A comparative study, Technical report, Equinox Corporation, 2002.

[22] **Segundo, MP., Silva, L., Bellon, ORP., Queirolo, CC.** Automatic face segmentation and facial landmark detection in range images. s.l. : IEEE Trans Syst Man Cybern Part B Cybern, 2010. pp. 1319–1330.

[23] **Chan, T.F., Vese, L.A.,.** *Active contours without edges.* s.l. : IEEE Transactions on Image Processing, 2001. pp. 266-277.

[24] **Filipe, S., Alexandre, L.A.** Thermal Infrared Face Segmentation: A New Pose Invariant Method. s.l. : 6th Iberian Conference on Pattern Recognition and Image Analysis, IbPRIA, Springer, 2013, pp. 632-639.

[25] **Shih, F., and Chuang, C.,.** *Automatic extraction of head and face boundaries and facial features.* s.l. : Information Sciences 158, 2004. pp. 117–130.

[26] **Wong, K., Lam, K., and Siu, W.,.** *An efficient algorithm for human face detection and facial feature extraction under different conditions.* s.l. : Pattern Recognition 34, 2001. Vol. 34, pp. 1993-2004.

[27] **Butakoff. C., Frangi, AF.** A Framework for Weighted Fusion of Multiple Statistical Models of Shape and Appearance. 28 s.l. : Pattern Analysis and Machine Intelligence, IEEE Transactions, 2006. pp. 1847-1857 .

[28] **Hall, P., Marshall, D., Martin, R.** *Merging and Splitting Eigenspace Models.* s.l. : IEEE Trans. Pattern Analysis and Machine, 2000. Vol. 22, pp. 1042-1049.

[29] **Dargham, J., Chekima, A., and Pandiyan, P.,.** *Skin Detection using Neural Networks.* s.l. : Proceedings of the Second International Conference on Artificial Intelligence in Engineering & Technology, 2004. pp. 371-376.

[30] **Feng, Xiao-yi.** *Eyes Location by Neural Network-Based Face Segmentation .* s.l. : IJCSNS International Journal of Computer Science and Network Security, 2006.

[31] **Xiao, Y., and Yan, H.,.** *Facial Feature Location with Delaunay Triangulation/Voronoi Diagram Calculation.* s.l. : Australian Computer Society, Inc. The Pan-Sydney Area Workshop on Visual Information Processing, 2001. pp. 103-108.

[32] **Filipe, S., Alexandre, L.A.,.** *Improving Face Segmentation in Thermograms Using Image Signatures.* Sao Paulo, Brazil : Springer, 15th Iberoamerican Congress on Pattern Recognition, 2010. pp. 402-409.