SHORT PAPER

Algorithms for invariant long-wave infrared face segmentation: evaluation and comparison

Sílvio Filipe · Luís A. Alexandre

Abstract This paper presents two methods for automatic segmentation of images of faces captured in long wavelength infrared, allowing a wide range of face rotations, expressions and artifacts (such as glasses and hats). We also present the validation of segmentation results using a recognition method to show the impact of the segmentation accuracy on the recognition. The paper presents two different approaches (one aimed at real-time performance and the other at high accuracy) and compares their performance against three other previously published methods. The proposed approaches are based on statistical modeling of pixel intensities and active contour application, although several other image processing operations are also performed. Experiments were performed on a total of 893 test images from four public available databases. The obtained results improve on previous existing methods up to 29.5 % for the first measure error (E_1) and up to 34.7 % for the second measure (E_2), depending on the method and database. Regarding the computational time, our proposals improve up to 63.32 % when compared with the other proposals. We also present the validation of the various segmentation methods that are presented by applying a face recognition method.

Keywords Face segmentation · Thermal infrared · Human face segmentation · Image segmentation · Face recognition · Principal component analysis

S. Filipe (✉) · L. A. Alexandre
IT - Instituto de Telecomunicações, Department of Computer Science, University of Beira Interior, 6200-001 Covilhã, Portugal
e-mail: sfilipe@ubi.pt; lfbaa@di.ubi.pt

1 Introduction

During the last decades, there has been a great deal of research in the area of face recognition, especially in the visible spectrum. But recognition systems in the visible spectrum have problems dealing with the variations of light [14, 21]. To solve this problem, the proposed solutions are the use of 3D face recognition [2] or a combination of facial recognition in the visible and Infrared (IR) spectrum [1, 16].

The growing concern over security has led to interest in the development of more robust methods, giving rise to face recognition only in the infrared, since the long wavelength infrared (LWIR) recognition is not affected by variations of light.

Segmentation is more demanding than the simple face detection, since it not only points to the location of the face, but must also describe its shape. A good segmentation system can improve the recognition rates for most recognition methods, allowing the use of the shape of the face in the recognition process (see Fig. 1) [18, 22]. The goal of [18] is to define and address the issues associated with incorporating image segmentation into an object recognition framework. In [22], we see that the authors improved the results of face recognition by developing a segmentation method.

Contrary to the visible wavelength, where there are numerous methods for accomplishing this task (based on geometry [5], color [25], etc.,) in the LWIR there is a lack of proposals to improve the current status.

Figure 1 shows the general scheme for a recognition system. This scheme can be used for face recognition either in the visible or thermal wavelengths and can be used also for other recognition modalities, such as those that use iris images [20]. The recognition system has two parts:
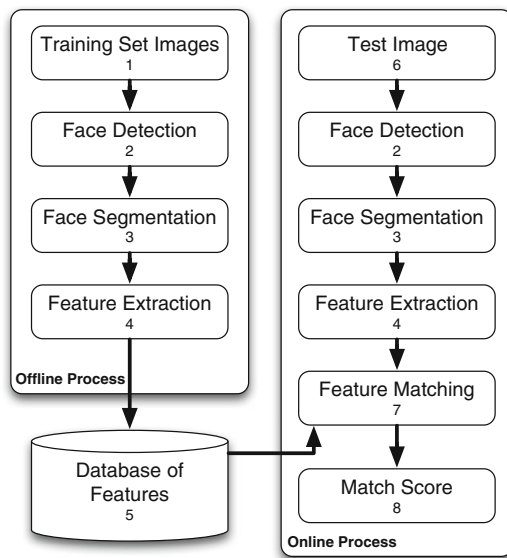
Fig. 1 Block diagram of a face recognition system

- *Offline process:* The training set images are captured by a camera. Face detection is done followed by a segmentation that obtains the face and features are extracted. These are stored in a database.
- *Online process:* Given an image, it is detected, segmented and features are extracted as in the offline process. These are compared against the ones stored in the database and a match score is produced.

Note that not all recognition algorithms use all of the above steps. Sometimes only detection is used and there is no segmentation step [32]; the inverse can also occur as the face detection methods can have problems in detecting faces that are not frontal [9, 19].

In the case of the two methods proposed in this paper, we do not assume a first stage of face detection prior to the application of the methods, but if such stage is used, no change is required in the proposed methods.

In the next sections, we present a short description of the available LWIR face segmentation methods (Sect. 2) and present our proposed methods (Sect. 3) In Sect. 4, we present the datasets used and experimental results, including a discussion. We end the paper in Sect. 5 with the conclusions.

2 Overview of face segmentation in thermal infrared images

A preprocessing step for many of face recognition methods, which can lead to failure if not done correctly, is the segmentation of the face.

Gyaourova et al. in [13] proposed a method based on an elliptical mask that is placed over the face image. The problem is that this approach will work only on frontal faces, centered and captured at the same distance (in order to have approximately the same size).

Pavlidis et. al. [19] achieve face segmentation through a Bayesian approach, fitting two normal distributions per class applying an adaptation of the EM algorithm. This algorithm accepts skin (s) and background (b) pixels from selected subregions of the training set where only one of those types is present, then produces four means (μ), four variances (σ^2) and two weights (ω) These values are obtained by algorithm 1. In the segmentation stage, for each pixel there is a prior distribution $\pi^{(t)}(\theta)$, for the tth iteration, where θ is the parameter of interest that takes two possible values (s and b), whether it is a skin ($\pi^{(t)}(s)$) or a background ($\pi^{(t)}(b) = 1 - \pi^{(t)}(s)$) pixel. Its initial prior probability is given by $\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$.

Algorithm 1 Expectation Maximization (EM) algorithm for determining the initial parameters of the skin distributions. This algorithm takes as input parameters the pixels considered as skin (x) and its size (N_s). A similar process is applied to determine the initial parameters of the background distributions.

Input: x, N_s
$k = 0$
$\forall_{i \in \{1,2\}} \omega_{s_i}^{(k)} = \frac{1}{2}$
$\forall_{i \in \{1,2\}} \mu_{s_i}^{(k)} = \frac{1}{N_s} \sum_{j=1}^{N_s} x_j$
$\forall_{i \in \{1,2\}} \sigma_{s_i}^{(k)} = \frac{1}{N_s} \sum_{j=1}^{N_s} (x_j - \mu_{s_i}^{(k)})^2$
repeat
    for $i = 1$ to 2 **do**
        for $j = 1$ to N_s **do**
$$z_{ij} = \frac{\omega_{s_i}^{(k)}(\sigma_{s_i}^{(k)})^{-1} exp\left\{ -\frac{1}{2(\sigma_{s_i}^{(k)})^2}(x_j - \mu_{s_i}^{(k)})^2 \right\}}{\sum_{t=1}^{2} \omega_{s_t}^{(k)}(\sigma_{s_t}^{(k)})^{-1} exp\left\{ -\frac{1}{2(\sigma_{s_t}^{(k)})^2}(x_j - \mu_{s_i}^{(k)})^2 \right\}}$$
        end for
        $\omega_{s_i}^{(k+1)} = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)}}{N_s}$
        $\mu_{s_i}^{(k+1)} = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)} x_j}{N_s \omega_{s_i}^{(k+1)}}$
        $(\sigma_{s_i}^{(k+1)})^2 = \frac{\sum_{j=1}^{N_s} z_{ij}^{(k)} (x_j - \mu_{s_i}^{(k+1)})^2}{N_s \mu_{s_i}^{(k+1)}}$
    end for
    $k = k + 1$
until $|\omega_{s_i}^{(k-1)} - \omega_{s_i}^{(k)} \leq \epsilon|$
return $\omega_s, \mu_{s_1}, \mu_{s_2}, \sigma_{s_1}^2, \sigma_{s_2}^2$

The input pixel value x_t has a conditional distribution $f(x_t \mid \theta)$. If the particular pixel is skin, we have $f(x_t|s) = \sum_{i=1}^{2} \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$, where $\mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$ is the normal distribution with mean μ_{s_i} and variance $\sigma_{s\_i}^2$, and where ω_{s_2} is given by $1 - \omega_{s\_1}$.

Based on algorithm 1, we obtained the pixel intensity distributions shown in Fig. 2, where dashed lines represent the estimated distributions for skin pixels and dashed point

are used for the background. The solid lines show the pixel intensity distributions for the training images cropped from the four databases used in this paper (presented in Sect. 4.1). The choppy distributions in Fig. 2d are caused by the fact that images at the Florida State University (FSU) database only contain around 70 different values.

Some of the segmented images obtained using this method are presented in the sixth and seventh rows of Fig. 10.

More recently, Cho et al. in [9] presented a method for segmentation of the face in IR images based on contours and morphological operations. The edge detector used is the Sobel edge detector, where only the largest contour is used which is considered to be the best to describe the face. After that, they apply morphological operations to the contour in this area, to connect the open contours and remove small areas. Rows 4 and 5 of Fig. 10 show some segmented images using this method.

In [10], an extension of the method in [19] is presented. The extension consists in closing image regions that have been left with holes, based on edge detection and morphological operations. In that paper, the method is not a method

for face segmentation, but skin segmentation. The big difference between the skin segmentation and face segmentation relates to the fact that the neck is included or not in the segmentation, respectively. Because of this difference, we chose not to include the results of that method in our article, since now the segmentation masks (shown in the second and third rows of Fig. 10) do not include the neck.

3 Proposed methods

After we have evaluated the methods [19] and [9], we saw that it was possible to overcome some shortcomings of these methods, to improve their results.

Regarding the method in [9], we found that it frequently included the background as face pixels given that the applied morphological operations could leak into the background when the face border was not properly established.

The method in [19] is based on the models of skin (and not just face) and background pixel intensities. This resulted in including clothes pixels as skin pixels and also
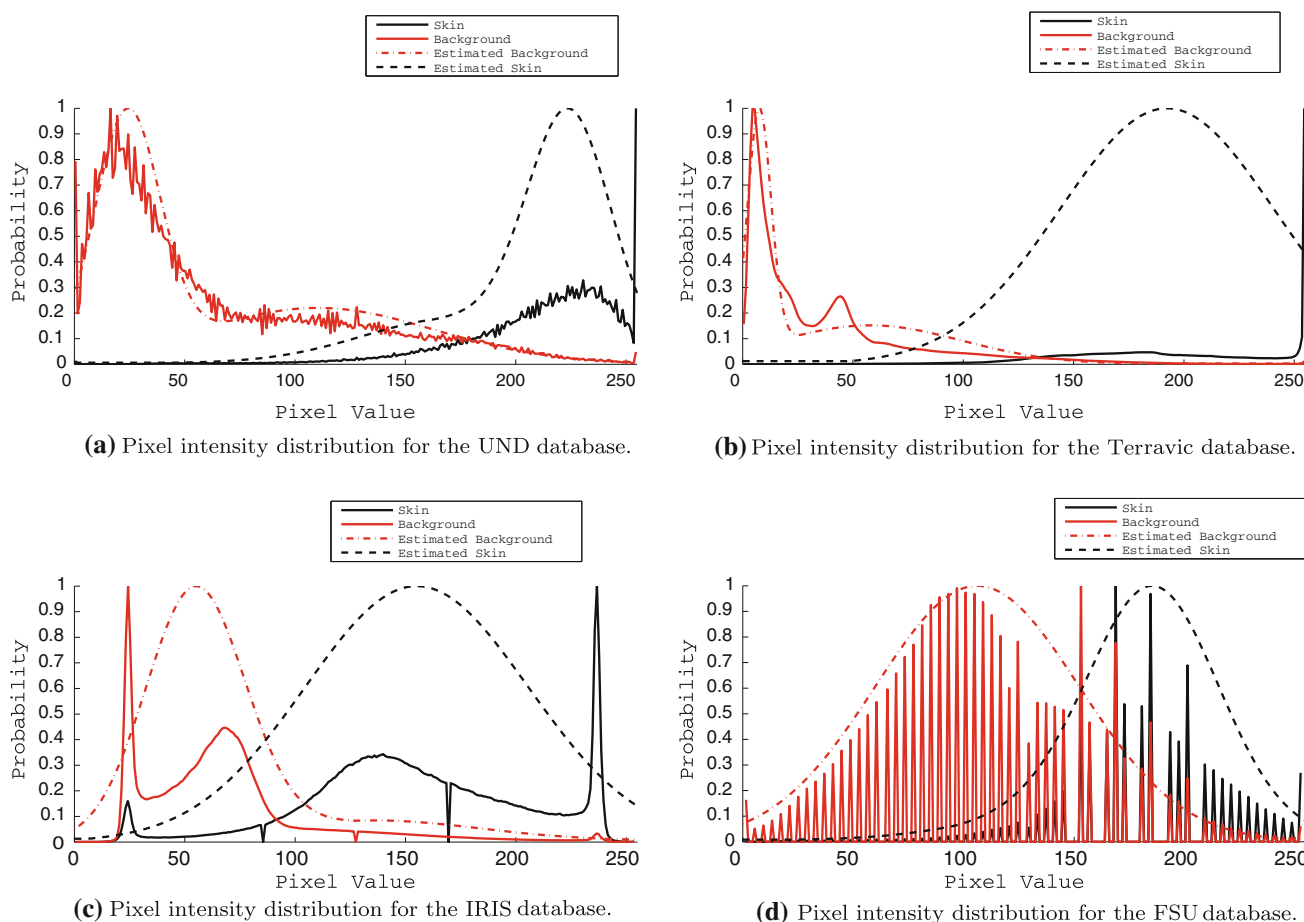

(a) Pixel intensity distribution for the UND database.

(b) Pixel intensity distribution for the Terravic database.

(c) Pixel intensity distribution for the IRIS database.

(d) Pixel intensity distribution for the FSU database.

Fig. 2 Face pixel intensity distribution (graphs to the *right*) and background (graphs to the *left*) for the four databases used

ignoring some skin pixels considering them to be background.

We designed two methods that are able to overcome the main problems identified with the existing face segmentation methods.

Method 1 was conceived to be simple and fast such that it could be used in real-time applications (see Fig. 3a). The idea is simply to look for the hottest (higher gray scale value) pixels that lie inside a rectangular region of interest (RROI). This RROI is obtained using image signatures, as described in Sect. 3.1. The threshold used to detect the hottest pixels is adaptive and obtained using the pixel distributions of the training set images, as described below in Sect. 3.3.

The second method we propose in this paper, which we call method 2, was designed to give more importance to accuracy than to speed. It starts by extracting the largest ellipse that fits into the RROI (see Fig. 4a). This ellipse is used as the first iteration of method [7]. To complete this method, we apply the operations described below as the face pixel identification from binary image (FPIBI).

The result of the application of methods 1 and 2 to the images in the first row of Fig. 10 is presented in the same figure in the last four rows.

In the following subsections, we will describe the steps used in both proposed methods.

3.1 Rectangular region of interest (RROI)

An interesting operator would give the RROI that contains the face. This would avoid the problems caused by the clothes since, as the body warms it, clothes have temperatures similar to the face. This can hinder may
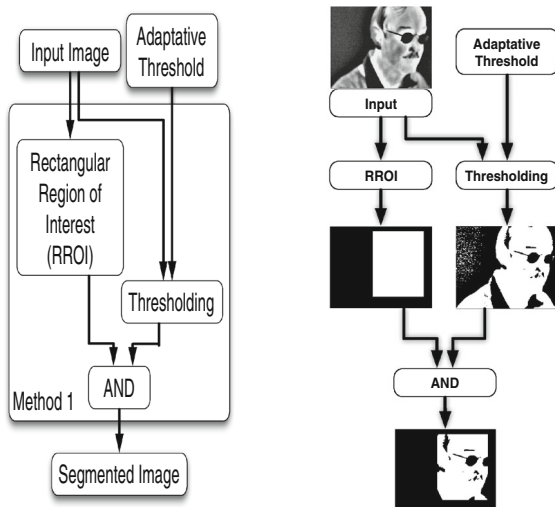
(a) Block diagram of method 1. **(b)** Illustration of method 1.

Fig. 3 Block diagram and illustration of the segmentation method 1
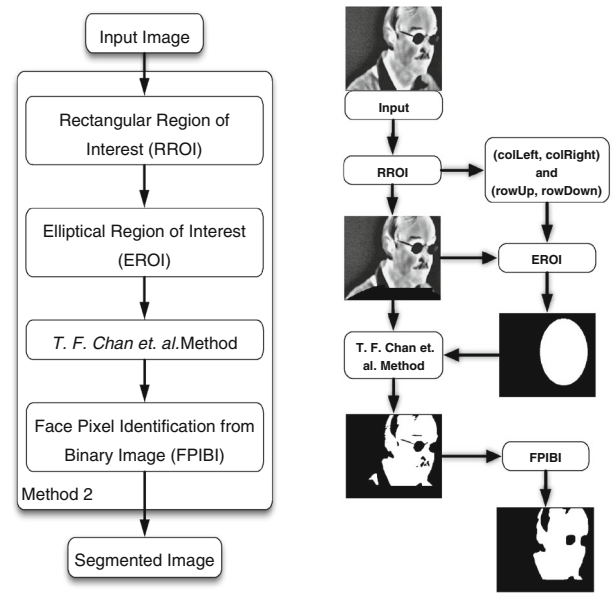
(a) Block diagram of method 2. **(b)** Illustration of method 2.

Fig. 4 Block diagram and illustration of the segmentation method 2

difficult the pixel intensity-based segmentation approaches.

To obtain the RROI, we will analyze the vertical and horizontal image signatures. These are 1D vectors that contain the sum of the intensity of the pixels along the columns and rows, respectively:

$$sigV(c) = \sum_{r=1}^{R} I(c,r) \tag{1}$$

$$sigH(r) = \sum_{c=1}^{C} I(c,r) \tag{2}$$

where c and r are the indexes of column and row for image I of dimension $C \times R$.

The first step to obtain the RROI is now described (see Fig. 5). We start by analyzing the vertical signature. The signal in Fig. 5b represents the vertical signature of Fig. 5a. This signal has several high-frequency oscillations that will appear in its derivative. This can be avoided by smoothing it with a 1D Gaussian filter (we used the one in Fig. 5c). The standard deviation of the Gaussian filter is $\sigma = 0.05 \times C$. This value was obtained by studying the influence of different values of σ in the training set images. The result of the convolution is in Fig. 5d and its derivative is in Fig. 5e. The following step consists in determining the extrema of this signal: in Fig. 5a, e we marked the maximum with the left dash line (*colLeft*) and the minimum with a right dash line (*colRight*). The two lines indicate the location of a large variation in image intensity that we identify with the sides of the face.
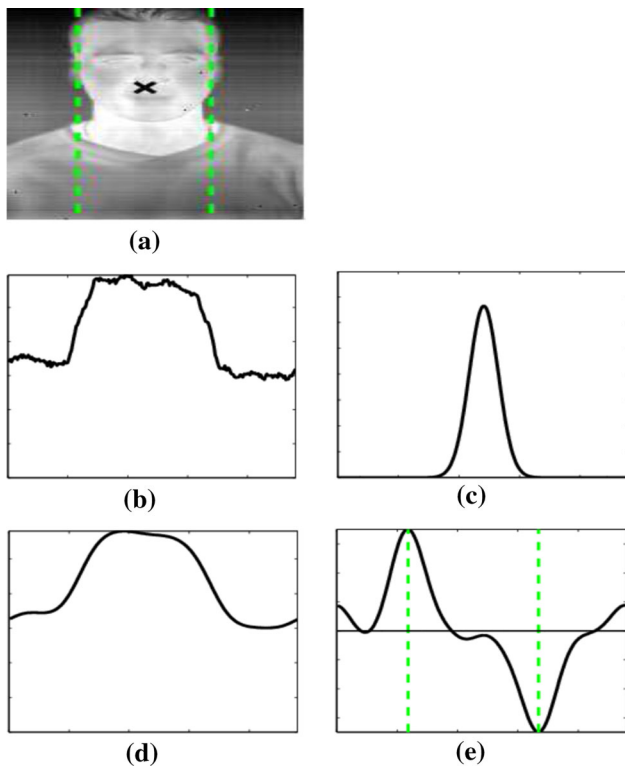
Fig. 5 Vertical signature analysis. Figure **a** is the original image. Figure **b** is its vertical signature. Figure **c** is the Gaussian filter used in the convolution. Figure **d** is result of applying this filter to **b**. Figure **e** is the derivative of figure **d**
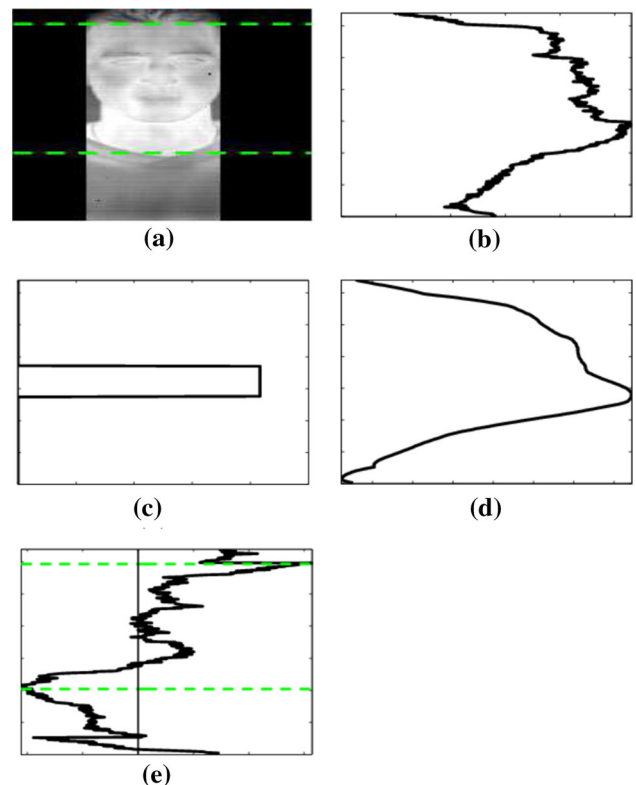

Fig. 6 Horizontal signature analysis. Figure **a** is the result of vertical signature and input image to the horizontal signature. Figure **b** is its horizontal signature. Figure **c** is the filter used in the convolution. Figure **d** is result of applying this filter to **b**. Figure **e** is the derivative of figure **d**

The next step in defining the RROI is the analysis of the horizontal signature to obtain the upper bound on the face (*rowUp*). For this, we only consider the part of the image between the two extrema detected in the vertical signature analysis. This removes the shoulders of the subjects and overcomes one of the issues that was causing problems in the previous approaches. The process used in the analysis of the horizontal signature (see Fig. 6) is similar to the one used to analyze the vertical signature. The main difference is the filter used: in this case, its width is $0.15 \times R$. The shape and size of the filter were selected to remove sudden variations that would appear in the signal when the subjects are wearing glasses or have a cold nose. Next, we used a process similar to the one for the vertical signature to obtain the extrema of the smoothed signal. Finally, the upper bound of the face (*rowUp*) is given by the maximum of horizontal signature and is represented by the upper dash line in Fig. 6a.

The delimitation of the lower face (*rowDown*) is given by fitting a parabola to the contours of the shoulders or the chin. Knowing that a person's shoulders are always at the bottom of the image, we analyzed only the region between $[\frac{2}{3} \times R, R]$ (where R is the number of rows of the image). In this region, a linear reduction in the number of colors was

done to enhance the chin, neck and shoulders regions and remove certain types of background noise (shown in Fig. 7c, d). After reducing the number of colors, we apply a Gaussian blur to smooth the abrupt changes in the regions presented (shown in Fig. 7e, f). The parameters used in the Gaussian filter are $\sigma = 2.5$ and the size is 25×25. We used the Canny edge detector [6] to obtain the points used to adjust our parabola. The parameters used in the Canny edge detector are $\sigma = 1.0$, low threshold = 0.2×255 and high threshold = 0.7×255. These were chosen to eliminate the lower edges (such as temperature variations on clothing or face, see Fig. 7g, h).

The result obtained by the Canny edge detector is used to fit a second-order function (parabola) to obtain the parameters a, b and c:

$$f(x) = ax^2 + bx + c \tag{3}$$

To find the parameters of the parabola (a, b and c), defined by Eq. 3, which best describes the curvature of the shoulders (figures in the first column of the Fig. 7) or the chin when the shoulders are not detected in the contours (figures in the second column of the Fig. 7), we use the least-squares method. This approach is standard to obtain
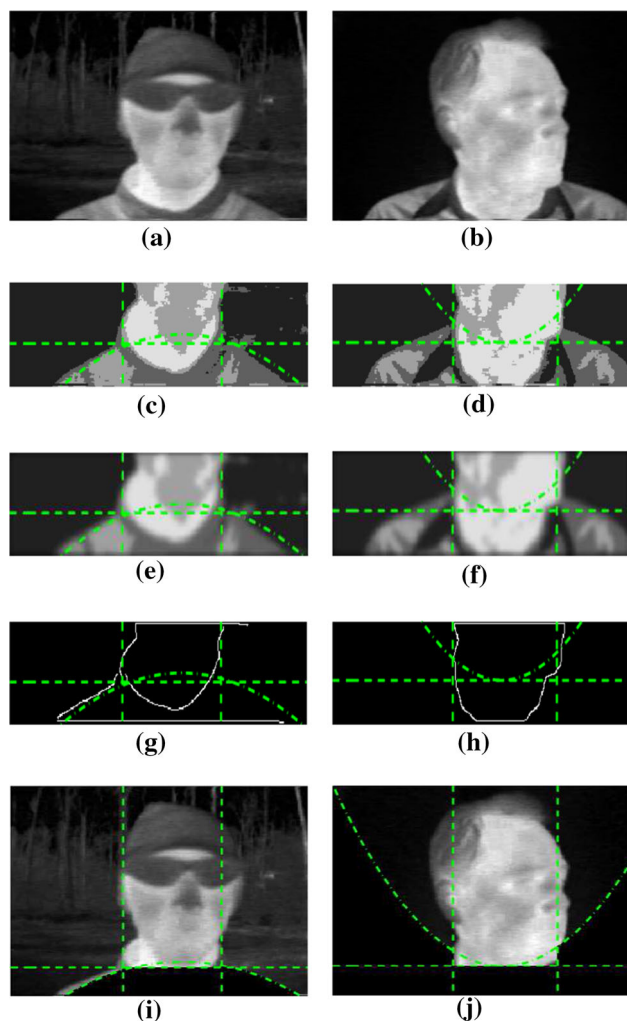
Fig. 7 Process for fitting a parabola. Figures **a** and **b** are the input images. Figures **c** and **d** are images with reduced number of colors. Figures **e** and **f** are images with Gaussian blur. Figures **g** and **h** are images with detected contours. Figures **i** and **j** contain the resulting parabola and the region below it is marked as background (between *colLeft* and *colRight* the delimitation of the background is given by *rowDown*)

an approximate solution of over-determined systems, i.e., sets of equations in which there are more equations than unknowns. The best fit in the least-squares sense minimizes the sum S of squared residuals:

$$S = \sum_{i=1}^{n} r_i^2, \tag{4}$$

where the residual (r_i) is the difference between an observed value and the fitted value provided by a model:

$$r_i = y_i - f(x_i). \tag{5}$$

rowDown is given by the mean value between *f(colLeft)* (*colLeft* is the left vertical line) and *f(colRight)* (*colRight* is the right vertical line) when $a > 0$ and when $a < 0$

rowDown is given by the 'vertex' of the parabola (maximum value of the parabola). In the images of Fig. 7 the *rowDown* is indicated by a horizontal line.

3.2 Elliptical region of interest (EROI)

The idea of defining an ellipse to enclosure the face is appealing, since a face has approximately the shape of an ellipse. An example is the work presented in [13], where the segmentation approach uses such an ellipse. We will also use an ellipse to improve the RROI around the face and to initialize the mask used in the method [7] (discussed in Sect. 3.4)

The ellipse will be defined inside the previously obtained RROI. We start by finding the center of the face, which we will use as the center of the ellipse. To determine this center point (*colCFace*, *rowCFace*), the cross on the image in Fig. 5a, we use the extrema obtained while searching for the RROI.

Algorithm 2 Algorithm for the Elliptical Region of Interest (EROI).

Input: $(colLeft, rowUp, colCFace, rowCFace)$
Output: $\{(X_0, Y_0), ..., (X_{2\pi}, Y_{2\pi})\}$
$diffCol = |colLeft - colCFace|$
$diffRow = |rowUp - rowCFace|$
for $\theta = 0$ to 2π **do**
   $X(\theta) = colCFace + (diffCol * cos(\theta))$
   $Y(\theta) = rowCFace + (diffRow * sin(\theta))$
end for

Then, using algorithm 2, we can obtain the $\{(X(0), Y(0)), ..., (X(2\pi), Y(2\pi))\}$ points of the ellipse. The algorithm receives the coordinates of the face center, (*colCFace*, *rowCFace*) and the coordinates of the left upper corner of the RROI, (*colLeft*, *rowUp*). These points are used for obtaining the distance from the center of the face to the left side of the RROI (which is denoted by *a*) and the distance from the center to the top of the RROI (denoted by *b*). *a* and *b* are used to convert the polar coordinates of the points that belong to the ellipse to Cartesian coordinates as $\{(X(0), Y(0)), ..., (X(2\pi), Y(2\pi))\}$.

3.3 Adaptative threshold

We will use a threshold step in method 1. The threshold is adaptive in the sense that it depends on the training set distributions for each database. The goal of this threshold is to separate most of the face pixels from the background, and therefore it will be chosen to guarantee that most of the face pixels will be included, although some of background pixels might also be included.

First, we identify the point at which the distributions (solid lines in Fig. 2) for face and background pixels

intersect. The threshold value is chosen as half the pixel value identified. Other rules might also work, but this is a simple one that yielded good results on training set experiments.

3.4 Active contours without edges

Based on the Mumford and Shah [17] minimal partition functional, Chan and Vese [7, 29] proposed a level set model for active contours to detect objects whose boundaries are not necessarily defined by the gradient, as with the classical active contour.

The main motivation for the use of this type of algorithms is their excellent ability to segment objects present in images. We choose for this step the use of the active contours without edges [7]. In [12], the authors refer that this method achieves greater accuracy and robustness at the cost of a major reduction in speed. For this work, we imposed a restriction: the maximum number of iterations is now 200. This reduces the computational cost without visible accuracy loss, according to some training set experiments performed. As will be seen below, the use of this approach will not be slower than the other methods [9] and [19]. The processing time will depend on the type of initial contour and of its position: the further away the initial contour is from the face, the longer it will take to converge.

An example of using this algorithm can be seen in [23] where it is used to segment teeth and where we can see that the X-ray images have some similarities with the images of the LWIR.

The result of segmenting the images in the first row of Fig. 10 with this method is in rows 8 and 9 of the same figure. To apply the active contour, we define an initial boundary as a centered rectangle of 90 × 140 pixels. This size was obtained by averaging the face size of the images in the training set. Method 2 will also use an active contour, but with an elliptical initial boundary. As we mentioned previously, the initial contour will affect the processing time, hence the importance of choosing an initial contour with a shape similar to a face.

3.5 Face pixel identification from binary image (FPIBI)

The result from the application of the active contour is used to select the face pixels in binary images (see Fig. 8a). We want to identify the largest contour that contains the face center and consider all pixels inside this contour as face pixels with the exception of the pixels that belong to glasses (see Fig. 8).

We start by identifying the center of the face as explained in the RROI operation (cross in the Fig. 8a). After that, we apply a dilation followed by an erosion (an
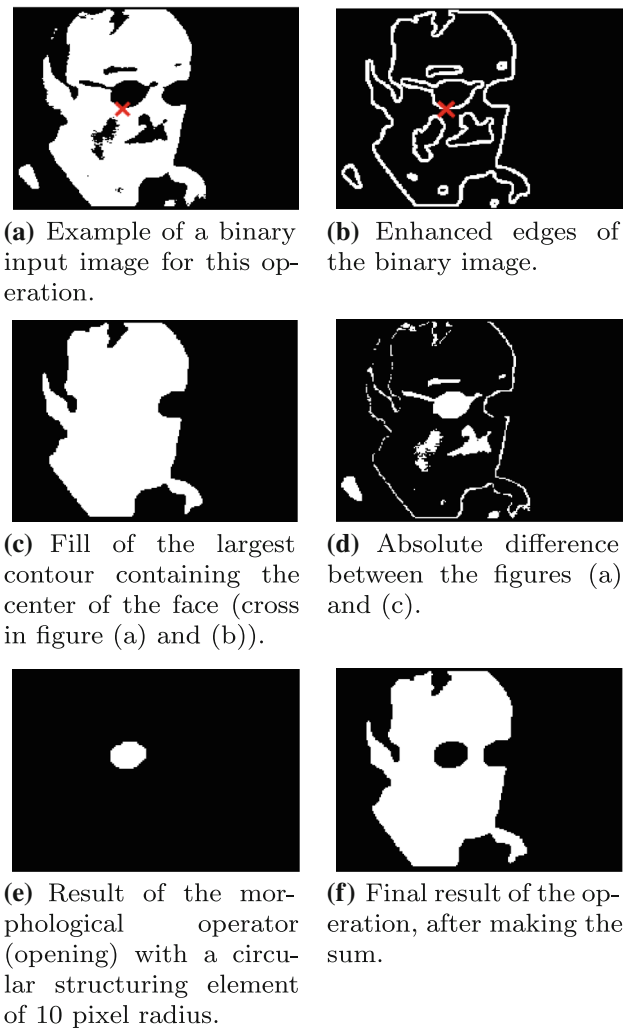

(a) Example of a binary input image for this operation.

(b) Enhanced edges of the binary image.

(c) Fill of the largest contour containing the center of the face (cross in figure (a) and (b)).

(d) Absolute difference between the figures (a) and (c).

(e) Result of the morphological operator (opening) with a circular structuring element of 10 pixel radius.

(f) Final result of the operation, after making the sum.

Fig. 8 Step-by-step example of the application of the FPIBI operation (see text for detailed description)

opening) using structural elements of sizes 3 × 3 and 2 × 2, respectively. These morphological operations are used to remove small areas and an edge map is obtained using the Canny edge detector [6]. The obtained edges are enhanced through a dilation with a structuring element of size 3 × 3 (see Fig. 8b). From these edges, we select the largest that contains the face center. We now assume that all the pixels inside this largest contour are face pixels (see Fig. 8c).

To remove glasses that may have been considered as being face in the previous step, we make the absolute difference between the image before the selection of the largest contour and the image that results from filling this largest contour (see Fig. 8d). With this difference, we will obtain the image regions that were altered by the filling. We apply an opening with a circular structuring element of 10 pixel radius (see Fig. 8e). Only the largest regions, such as the glasses, remain after the application of this

morphological operator. The resulting image is added to the one that results from filling the largest contour using the logical function (see Fig. 8f).

4 Experimental results

4.1 Datasets

The University of Notre Dame (UND) database is presented in [8, 11]. The 'Collection X1' of the UND database contains 2,293 LWIR frontal face IR images from 81 different subjects. The training set contains 159 images and the test set 163. Two images from this database are in the first row, columns 1 and 2 of Fig. 10.

The 'Dataset 04: Terravic Facial IR Database' is a subset of the object tracking and classification in and beyond the visible spectrum (OTCBVS) database [31]. This database contains 24,508 images of 20 different persons. It has different poses (rotations front, left, right), images captured indoor and outdoor, and images of people with glasses and hats. The training set has 235 images and the test set has 240. Two images from this database are in the first row, columns 3 and 4 of Fig. 10.

The 'Dataset 02: IRIS Thermal/Visible Face Database' is a subset of the OTCBVS database [30]. The database contains 4,228 images that were acquired in the Imaging, Robotics, and Intelligent Systems laboratory (University of Tennessee) (IRIS) with 11 images per rotation (images for each expression and illumination) yielding between 176 and 250 images per person. This database was acquired with different illuminations in the visible wavelength. These differences do not affect the LWIR; therefore, we ignore the different versions due to illuminations changes. The training and test sets have both 296 images. Two images from this database are in the first row, columns 5 and 6 of Fig. 10.

The FSU database contains 234 frontal IR images of ten different subjects, which were obtained at varying angles and facial expressions [26]. The training set contains 40 IR images (four per subject) and the test set 194. Two images from this database are in the first row, columns 7 and 8 of Fig. 10.

The test set images from all databases were segmented manually to create the test set ground truth (samples shown in row 2 of Fig. 10). Method [9] does not need a training set and the method [19] and ours use pixels information from manually segmented regions of the training set images. With that, these methods need an accurate segmentation of the training set.

Table 1 shows the percentage of face and background pixels present in the test sets used in this paper. These values are obtained based on the manually segmented

Table 1 Face and background pixel ratios to the total number of image pixels, for the different databases (values obtained in the used test sets)

	Pixel percentage	
	Face	Non-face
UND	12.13	87.87
Terravic	17.67	82.33
IRIS	26.75	73.25
FSU	50.53	49.47

images. We can see that the FSU database is the only that has more face than background pixels (the other databases have a face to background pixel ratio between 12.13 and 26.75 %).

A list with the names of the images used in the train and test sets, code and segmentation masks are available at: *hidden link for blind review*.

4.2 Evaluation

The requested task is quite simple: for each input image (as the ones shown in the first row of Fig. 10) a corresponding binary output (as those shown in the second row of the same figure) should be built, where the pixels that belong to the face and are noise-free should appear as white, while the remaining pixels are represented in black. The test set of the databases was used to measure pixel-by-pixel agreement between the binary maps produced by each of the algorithms (these maps are shown in Fig. 10, rows 4, 6, 8, 10 and 12) and the ground-truth data manually built a priori (see examples in row 2 of Fig. 10).

The classification error rate (E_1) of the algorithm on the input image I_i ($E_1(i)$) is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator) across the image:

$$E_1(i) = \frac{1}{C \times R} \sum_{c=1}^{C} \sum_{r=1}^{R} O(c,r) \otimes T(c,r) \tag{6}$$

where $O(c,r)$ and $T(c,r)$ are, respectively, pixels of the output and true class images. C and R are the number of columns and rows, respectively.

The classification error rate (E_1) of the algorithm is given by the average of the errors on the n test images $E_1(i)$:

$$E_1 = \frac{1}{n} \sum_{i=1}^{n} E_1(i) \tag{7}$$

The value of E_1 is in the [0,1] interval and 1 and 0 will be, respectively, the worst and best values.

The second error measure (E_2) aims to compensate the disproportion between the a priori probabilities of 'face'

and 'non-face' pixels in the images. The type-I and type-II error rates are given by the average between the false positive rate (FPR) and false negative rate (FNR):

$$E_2(i) = \frac{\text{FNR}}{2} + \frac{\text{FPR}}{2} \qquad (8)$$

where the FPR is given by:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \qquad (9)$$

and the FNR by:

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TP}}. \qquad (10)$$

where FN is the false negative, TN is the true negative, FP the false positive and TP is the true positive.

Similarly to the E_1 error rate, the final E_2 error rate is given by the average of the errors ($E_2(i)$) on the n test images:

$$E_2 = \frac{1}{n} \sum_{i=1}^{n} E_2(i) \qquad (11)$$

4.3 Experimental results and discussion

In this section, we present and discuss the experiments performed during this work.

Each of the methods presented in the paper was developed in Matlab R2009b and evaluated individually on an Intel Core 2 Q9300 (2.5 GHz), 4 Gb RAM (FSB 1066) and Fedora Core 11 operative system, so that there is no competition for access to the computer resources.

In each algorithm, we evaluate its accuracy, by measuring the errors E_1 and E_2, and its execution time.

The quantitative evaluation of the proposed methods is presented in Table 2. It contains the errors E_1 and E_2 of each algorithm. The error rates are also shown in the graphs of Fig. 9a, b to allow a quick comparison between each method and database.

Regarding the results for the UND database, method 2 improves the results between 3.3 and 31.6 %. The method [19] only analyzes the distribution of intensities of the pixels, so that when there is a region of clothes that has a temperature similar to the skin it is considered to be skin. In this database, the method [7] does not have better results because it is an iterative method with two stop conditions (maximum number of iterations and the absolute difference between iteration i and $i-1$ be less than 1×10^{-3}). It only reaches the maximum number of iterations, never stopping because of the other condition, causing the active contour to spread through the region where there are clothes. Our methods have similar error rates in this database. This is because most of the faces are centered on the image and do not have any type of rotation. The biggest
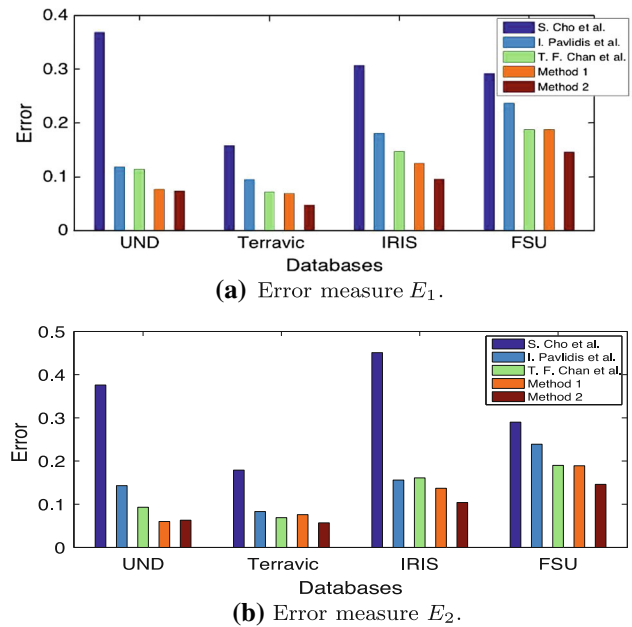

(a) Error measure E_1.

(b) Error measure E_2.

Fig. 9 Graphical representation of error measures E_1 and E_2, and execution time from Table 2

difference is the execution time, because method 1 has a much smaller execution time without losing quality in the segmentation.

The second part of Table 2 shows the results of the methods for segmenting the Terravic database, and the improvement of our methods range from 1.2 to 12.2 % (error measure E_2). Our methods obtained only minor improvements in this database, because there were a lower proportion of clothes in the images and the part that appeared in the images had a lower temperature than the face since part of the database was captured outside.

For the IRIS database, the results are presented in the third part of the Table 2. All methods had an increase in error rates for both measures of error (E_1 and E_2). In this database, the method in [19] considered many of the regions of hair and neck as part of the face. This is due to the existence of larger regions of face in the images which increased the detail of the hair and neck.

The method in [7] had large FNR because it considers parts of the face as background. In method 1, the error increased due to cuts made as a consequence of the analysis of the vertical and horizontal signatures, and also due to cuts in the chins made by the parabola. In method 2, the increase was not as sharp since the cuts were made solely by the fit of the parabola. Still, our methods have improved the segmentation results between 5.1 % (in measure E_1) to 34.7 % (in measure E_2).

The FSU database is the only database used where the number of face pixels is approximately equal to the number of background pixels. This resulted in four of the five methods
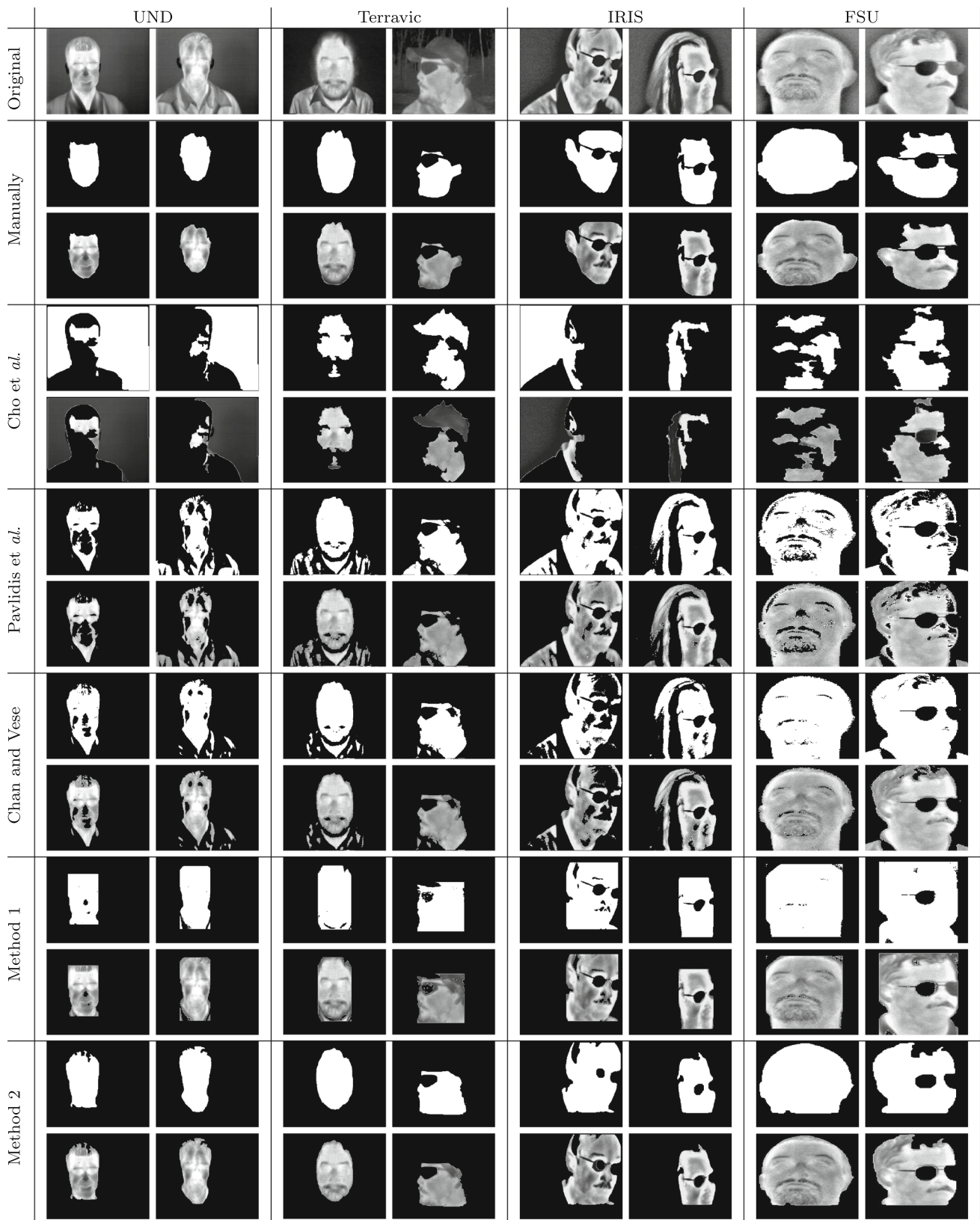
Fig. 10 Sample test set results for images from the UND database (columns 1 and 2), Terravic database (columns 3 and 4), IRIS database (columns 5 and 6) and FSU database (last two columns). Original images are in the first row and manual masks in the second. Approaches by [7, 9, 19], are in rows 3–5 respectively. The remaining rows have the results of ours methods 1 and 2, respectively

Table 2 Experiment results in all four databases

		Methods				
		Cho et al. [9]	Pavlidis et al. [19]	Chan and Vese [7]	Method 1	Method 2
UND	FNR	0.386	0.177	0.066	0.038	0.048
	FPR	0.366	0.109	0.121	0.082	0.078
	E_1	0.369	0.118	0.114	0.077	**0.074**
	E_2	0.376	0.143	0.093	**0.060**	0.063
	Time (s)	1.236	4.480	9.278	**0.039**	5.974
Terravic	FNR	0.210	0.066	0.065	0.086	0.071
	FPR	0.147	0.101	0.073	0.066	0.042
	E_1	0.158	0.095	0.072	0.069	**0.047**
	E_2	0.179	0.083	0.069	0.076	**0.057**
	Time (s)	1.237	4.554	10.361	**0.038**	7.401
IRIS	FNR	0.760	0.104	0.192	0.161	0.122
	FPR	0.142	0.209	0.131	0.112	0.087
	E_1	0.307	0.181	0.147	0.125	**0.096**
	E_2	0.451	0.156	0.161	0.137	**0.104**
	Time (s)	0.959	3.233	8.717	**0.032**	7.686
FSU	FNR	0.490	0.097	0.053	0.100	0.096
	FPR	0.091	0.381	0.327	0.278	0.197
	E_1	0.292	0.237	0.188	0.188	**0.146**
	E_2	0.290	0.239	0.190	0.189	**0.146**
	Time (s)	0.379	1.296	6.039	**0.025**	5.622

The first part presents the results for the UND database, the second of the Terravic database, the third of the IRIS database and the final part the results of the FSU database. The best (smallest) results in terms of error measures and execution time for each database are shown in bold

presented here increasing their FPR. Only the method in [9] had an FPR of less than 10 %, but otherwise has a FNR of 49 %. The increase in the FPR was due to the algorithms considering large regions of hair and neck as face pixels. The improvements made by our methods reflect the fact that the parabola cuts much of the neck. With this, our methods achieved an improvement of between 4.4 and 14.6 %.

In Fig. 10, we can see some examples of images of the test sets of the four databases (first row) used and what were the results of the segmentation methods in these images (rows 3–12). The second row of Fig. 10 shows the manually segmented images, which would be the optimal outcome for the result of a method. The fourth and fifth rows of Fig. 10 contain the result of the method in [9]. We observe that when the extracted contour is not closed, it assumes that much of the background is part of the face. When the extracted contour is closed, this method can find most of the pixels that are part of the face.

In the rows 6 and 7, the segmentation result of the method in [19] is shown. This method is based on a model of the distribution of the pixel intensities, which means that all regions that have a higher temperature (higher intensity

of pixels) are assumed as part of the face. Because the clothes are very close to the heat source (the body) they tend to have the same temperature as the body. Another problem with this approach appears when the facial skin is cold (due to people having been in a cold place, for instance), which makes it consider the coldest part of the face as background.

In this work, we also showed the use of a generic segmentation method based on active contours, which can be used in any type of images or objects that we want to target [7]. The results of this method are presented in rows 8 and 9 of Fig. 10 and we can see that it obtained good results, considering that it is a generic method. The problems with this method are similar to the method in [19], but the fact that we limited the maximum number of iterations to 200 meant that it did not include as many pixels belonging to the clothes as it otherwise would.

The results of our methods can be seen in rows 10–14 of Fig. 10. To try to solve the problems presented by the other methods, we defined steps strategically targeted to these problems. Looking at the results of method 1, shown on rows 10 and 11 of Fig. 10, we see that even using an extremely fast method we can solve much of the problem of clothes. The problem which remains is when parts of the face are cold, since these regions are rejected by the adaptive threshold.

To obtain a more accurate method, we had to increase the running time, and this lead to method 2. Through the combination of several operations that are included in this method, it can approximate quite well (rows 13 and 14 of Fig. 10) the desired results of manually segmented images (shown in the second and third rows of Fig. 10).

4.4 Validation

In this section, we validate the results obtained by the different segmentation methods presented. This validation involves the application of a method for face recognition.

Principal component analysis (PCA) is perhaps the most popular algorithm in the field [24, 27, 28] and it is a technique commonly used in dimensionality reduction in computer vision and particularly in face recognition. Principal component analysis techniques choose a linear projection that reduces the dimensionality while maximizing the scatter of all projected samples.

The *face space* is computed by taking a set of training observations, and finding the eigenfaces of this set. The training set of observations is given by the leave-one-out cross-validation (LOOCV) method [15], based on the manually segmented images. The image left out of the training set is used for comparison, and this image will change depending on the segmentation method used for validation. Thus, all the segmentation methods are

validated using the same eigenfaces. This is done, to force the recognition method to do the recognition based on face, and not on the clothing or other objects that the segmentation methods cloud identify as part of the face. The segmentation mask defines the region of the face to cut and we resize this region to 32×32. These steps are done before computing the eigenfaces.

To perform this validation, we will use three measures, which are the receiver operator characteristic (ROC) curve, the area under the ROC curve (AUC) and the decidability (DEC). The decidability index (Eq. 12) maximizes the distance between the distributions obtained for the two classical types of biometric comparisons: between signatures extracted from the same (*intra-class*) and different faces (*inter-class*).

$$\text{DEC} = \frac{|\mu intra - \mu inter|}{\sqrt{\frac{1}{2}(\sigma^2_{\text{intra}} + \sigma^2_{\text{inter}})}} \tag{12}$$

where μ_{intra} and μ_{inter} denote the means of the intra- and inter-class comparisons, σ^2_{intra} and σ^2_{inter} the respective standard deviations and the decidability can vary between $[0, \infty]$.

The obtained AUC and DEC are given in Table 3, while the ROC are presented in Fig. 11. In Table 3, the results presented in column Manually are the results obtained when only manual segmentation is used. These results are considered the best possible results for recognition using the PCA. We also added information on the number of times that the method achieved the best result in the recognition (*Wins* row) and the sums of scores (*Rank* rows), depending on the classification method in a given database. The scores assigned range from 5 to 1 points, where 5

points are assigned to the method that obtained the best result and 1 point to the worst.

Analyzing the recognition results obtained for the UND database, we can see that the AUC of our methods are very close (as well as can be seen in Fig. 11a) and that our best results shows an improvement of 4.3 % compared to previously proposed segmentation methods. Regarding the DEC, improvements are more significant, as can be seen since the distribution of intra-class is further away from the inter-class.

In the Terravic and IRIS databases, we can see that our method 2 produces results very similar to the results obtained in the recognition using the manual segmentation, with only a difference of 2.4 and 2.0 % relative to the AUC as shown in Fig. 11b, c respectively. For the DEC, the results of our methods were not as close to the ideal value (obtained by manual segmentation) as the AUC. Still, in Fig. 11c we see that the graph of method 2, when the FPR varies between [0.3, 0.5], overlaps with the graph of the manual segmentation and the method 1 is very close to the curve of the manually segmentation. Compared with other (previously published) segmentation methods, we achieved a significant improvement in both databases.

For the FSU database, the recognition based on our segmentation methods does not have the best results, as happened with the previous databases. With regard to ROC (shown in Fig. 11d), we see that the graph of our method 2 approximates the graph of the best method (Pavlidis et al. [19]) after reaching FPR = 0.3. The fact that our methods did not obtain better results in this database relates to the filter used in the analysis of horizontal signatures (shown in Fig. 6c). This is the only database where the face takes up almost the entire image, making the size of the filter

Table 3 Recognition results in all four databases

		Methods					
		Manually	Cho et al. [9]	Pavlidis et al. [19]	Chan and Vese [7]	Method 1	Method 2
UND	AUC	0.802	0.520	0.520	0.513	**0.556**	0.553
	DEC	1.155	0.079	0.074	0.057	**0.233**	0.169
Terravic	AUC	0.708	0.649	0.527	0.574	0.626	**0.684**
	DEC	0.831	0.537	0.127	0.285	0.514	**0.685**
IRIS	AUC	0.630	0.539	0.569	0.569	0.605	**0.610**
	DEC	0.504	0.124	0.266	0.230	0.328	**0.345**
FSU	AUC	0.727	0.588	**0.667**	0.636	0.600	0.642
	DEC	0.897	0.289	**0.613**	0.499	0.369	0.485
Wins			0	1	0	1	**2**
Rank	AUC	8	11	11	14	**18**	
	DEC	9	11	9	14	**17**	

The best (bigger) results in terms of AUC and decidability (DEC) for each database are shown in bold. The *Wins* row refers to the number of times the method has the best result. The *Rank* rows refer to the sum of the scores assigned to each of the methods, since the best (bold) is assigned with 5 points, the second 4 points, the third 3 points, the fourth 2 point and the worst 1 points

(a) ROC curve for the UND dataset.

(b) ROC curve for the Terravic dataset.

(c) ROC curve for the IRIS dataset.

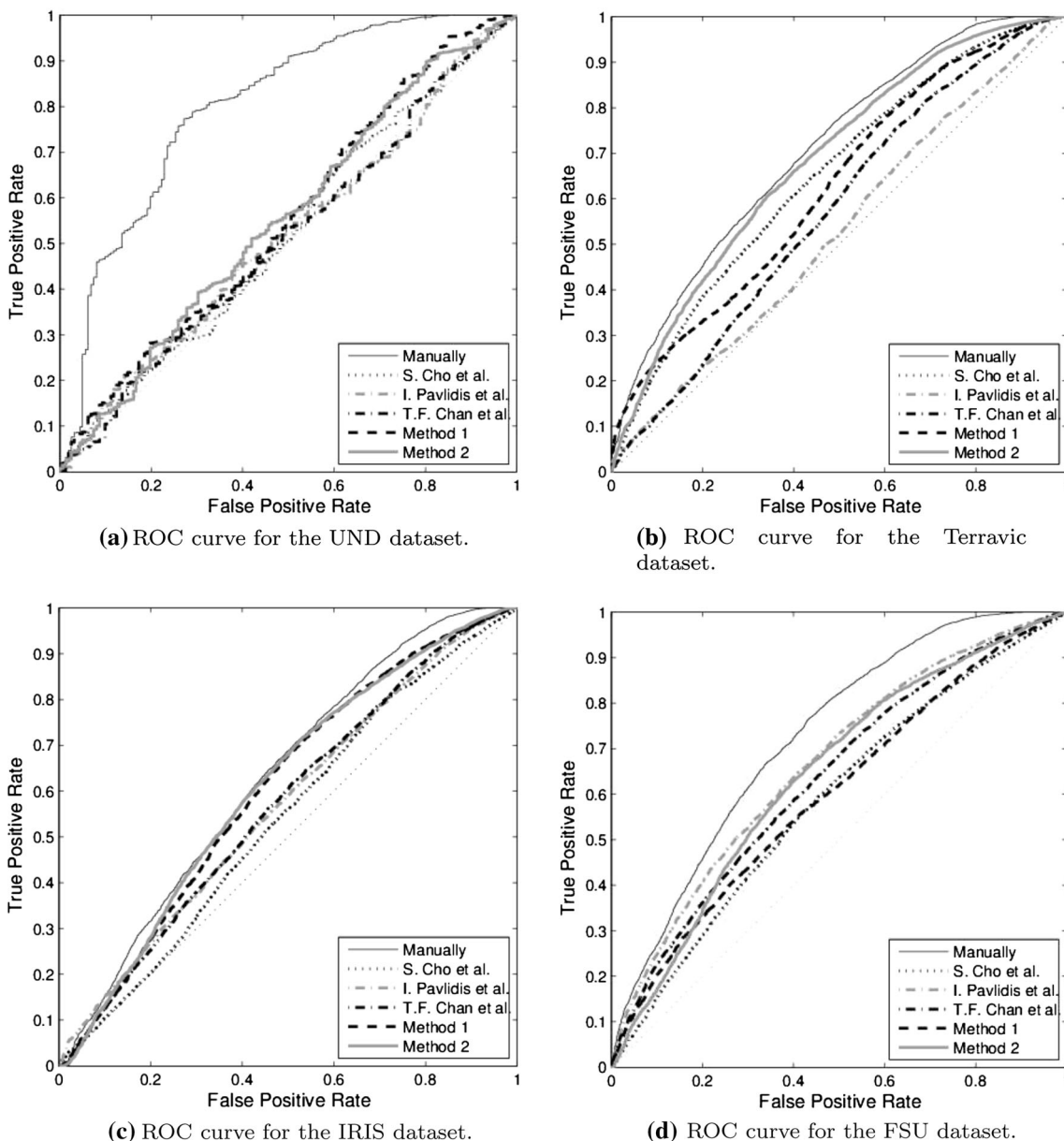(d) ROC curve for the FSU dataset.

Fig. 11 ROC curves for all databases, to validate the presented segmentation methods based on the recognition using PCA

(dependent on image size) small when the faces have glasses that occupy a large part of the image. The images with glasses take up about 1/4 of the face. For these images, our methods will strip away the forehead, and recognition is made only with the region that is between the eyes and chin.

Looking now to the results with a more global view (through the number of *Wins* and *Rank*), we note that our methods had the best ranks, even if they did not achieve the best AUC and DEC results for all databases, because when they were not the best method they were relatively close to the best.

4.5 Validation with artifacts separation

To improve the validation of our methods, we will present here the results of recognition (shown in Table 4) while separating the images containing artifacts from those without them. We only present the results for two databases (Terravic and IRIS), because the UND database only contains frontal faces without artifacts and the FSU the database contains few images with artifacts about 7 % of the total number of images). The images considered with artifacts are images where the people have glasses, hats, caps.

Table 4 Recognition results for the two databases (Terravic and IRIS) where there is a significant number of images with and without artifacts

		Terravic		IRIS		Rank
		With	Without	With	Without	
Manually	AUC	0.711	0.756	0.695	0.712	–
	DEC	0.844	1.061	0.744	0.778	–
Cho et al. [9]	AUC	0.594	0.711	0.542	0.547	8
	DEC	0.360	0.819	0.155	0.173	8
Pavlidis et al. [19]	AUC	0.560	0.501	0.572	0.588	7
	DEC	0.258	0.075	0.291	0.354	8
Chan and Vese [7]	AUC	0.571	0.569	0.600	0.583	9
	DEC	0.320	0.298	0.265	0.312	8
Method 1	AUC	0.622	0.640	0.597	0.626	14
	DEC	0.496	0.538	0.369	**0.445**	16
Method 2	AUC	**0.673**	**0.721**	**0.639**	**0.628**	20
	DEC	**0.671**	**0.827**	**0.456**	**0.445**	20

The best (bigger) results in terms of AUC and DEC for each database are shown in bold. The *Rank* columns refer to the sum of the scores assigned to each methods, similar to Table 3

As seen in Table 4, the method 2 achieved the best results (for both measures) in two databases. Only in the FSU database without artifacts, the method 1 has obtained similar results for AUC and the same value for the DEC. Analyzing the relationship between the results of recognition of our best method (method 2) and the previously proposed methods, we can see that there was an improvement of between 1 and 22 % for AUC and 0.008–0.752 in the DEC. These variations were all obtained using the Terravic database without artifacts, and the variations for the other sets are in these ranges.

5 Conclusion

This paper aimed at improving the state of the art in LWIR face segmentation. It contains a brief summary of the best existing methods, the proposal of two new methods that would perform well regardless of face pose, rotation and artifacts, expression, and an extensive evaluation of these methods under four different publicly available databases (730 training and 893 test images).

The segmentation evaluations were made taking into consideration two error measures that enable a more in depth analysis of the results: while E_1 is the usual error measure, E_2 takes into account the different number of points in each class (it is a balanced error measure).

The proposed methods were designed with two different goals: method 1 is aimed at real-time performance and it is the fastest of all methods in all databases, in some cases by

a very large amount. It does this without compromising the accuracy: it is equal or better than all the previous methods in both error measures with only one exception (Terravic database against the method in [7]). In terms of recognition, this method achieved good results, as seen in Tables 3, 4, where we validate the segmentation methods by applying a recognition method. Using method 1, we believe it is possible to perform face recognition in real-time using LWIR images.

Method 2 was developed to be accurate. It does this quite well, since it is the best in both error measures with the exception of the UND database, with improvements of up to 29.5 % according to E_1 and 34.7 % according to E_2, depending on the database. Of all the segmentation methods presented here, this is the one that is closer to the results of manual segmentation. This method, besides being able to solve the problem of including the clothes as part of the face, allows us to have the approximate shape of the face, which ultimately can be used by recognition methods. Nonetheless, it is not the slowest of all (this is the method in [7]), but we would advise the use of this method mainly on offline tasks. Based on two validations performed in Sects. 4.4 and 4.5, we can see that there was a significant improvement in the recognition results when compared to the results obtained using the segmentation by all other methods. The recognition results are close to results obtained for the manually segmented images, even, as Table 4 shows, when the recognition in done on images with artifacts.

References

1. Akhloufi M, Bendada A, Batsale J (2008) State of the art in infrared face recognition. Quant Infrared Thermogr J 5(1):3–26
2. Bowyer K, Chang K, Flynn P (2004) A survey of approaches to three-dimensional face recognition. 17th international conference on pattern recognition (ICPR 2004) 1:358–361
3. Buddharaju P, Pavlidis I (2007) Multispectral face recognition: fusion of visual imagery with physiological information. In: Face biometrics for personal identification, chap. 7, Springer, Berlin, Heidelberg, pp 91–108

4. Buddharaju P, Pavlidis I, Tsiamyrtzis P (2005) Physiology-based face recognition. In: IEEE conference on advanced video and signal based surveillance, pp 354–359

5. Butakoff C, Frangi AF (2009) Multi-view face segmentation using fusion of statistical shape and appearance models. Comput Vis Image Underst 114(3):311–321

6. Canny J (1986) A computational approach to edge detection. IEEE Trans PAMI 8(6):628–633

7. Chan TF, Vese LA (2001) Active contours without edges. IEEE Trans Image Process 10(2):266–277

8. Chen X, Flynn P, Bowyer K (2005) IR and visible light face recognition. Comput Vis Image Underst 99:332–358

9. Cho S, Wang L, Ong W (2009) Thermal imprint feature analysis for face recognition. IEEE Int Symp Ind Electron pp 1875–1880

10. Filipe S, Alexandre LA (2010) Improving face segmentation in thermograms using image signatures. In: Proceedings of 15th Iberoamerican congress on pattern recognition. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp 402–409

11. Flynn P, Bowyer K, Phillips P (2003) Assessment of time dependency in face recognition: an initial study. In: Proceedings of 4th international conference audio- and video-based biometric person authentication. Lecture Notes in Computer Science, vol 2688. Springer, Berlin, Heidelberg, pp 44–51

12. Freedman D, Radke RJ, Lovelock DM, Chen GTY (2005) Model-based segmentation of medical imagery by matching distributions. IEEE Trans Med Imaging 24(3):281–292

13. Gyaourova A, Bebis G, Pavlidis I (2004) Fusion of infrared and visible images for face recognition. In: Proceedings of 8th European conference on computer vision. Lecture Notes in Computer Science, vol 3024. Springer, Berlin, Heidelberg, pp 456–468

14. Jain A, Flynn P, Ross A (2007) Handbook of Biometrics. Springer, New York

15. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence-volume 2 (IJCAI'95). San Francisco, CA, USA, p 1137–1143

16. Kong S, Heo J, Abidi B, Paik J, Abidi M (2005) Recent advances in visual and infrared face recognition: a review. Comput Vis Image Underst 97(1):103–135

17. Mumford D, Shah J (1989) Optimal approximations by piecewise smooth functions and associated variational problems. Comm Pure Appl Math 42:577–685

18. Pantofaru C (2008) Studies in using image segmentation to improve object recognition. Ph.D. thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA

19. Pavlidis I, Tsiamyrtzis P, Manohar C, Buddharaju P (2006) Biometrics: face recognition in thermal infrared. In: Biomedical engineering handbook, 3rd edn. chap 29, CRC Press, p 1–15

20. Proença H, Alexandre L (2010) Iris recognition: analysis of the error rates regarding the accuracy of the segmentation stage. Image Vis Comput 28:202–206

21. Ross A, Nandakumar K, Jain A (2006) Handbook of multibiometrics (international series on biometrics). Springer, New York

22. Segundo MP, Silva L, Bellon ORP, Queirolo CC (2010) Automatic face segmentation and facial landmark detection in range images. IEEE Trans Syst Man Cybern Part B Cybern 40(5): 1319–1330

23. Shah S, Abaza A, Ross A, Ammar H (2006) Automatic tooth segmentation using active contour without edges. In: 2006 biometrics symposium: special session on research at the biometric consortium conference, Baltimore, MD, p 1–6

24. Sirovich L, Kirby M (1987) Low-dimensional procedure for the characterization of human faces. J Opt Soc Am A 4(3):519–524

25. Sobottka K, Pitas I (1996) Segmentation and tracking of faces in color images. In: Second international conference on automatic face and gesture recognition, Killington, pp 236–241

26. Srivastava A, Liu X (2003) Statistical hypothesis pruning for identifying faces from infrared images. Image Vis Comput 21:651–661

27. Turk M, Pentland A (1991) Eigenfaces for recognition. J Cognit Neurosci 3:71–86

28. Turk M, Pentland A (1991) Face recognition using eigenfaces. In: IEEE conference on computer vision and pattern recognition (CVPR'91), Maui, HI, USA, pp 586–591

29. Vese LA, Chan TF (2002) A multiphase level set framework for image segmentation using the Mumford and Shah model. Int J Comput Vis 50(3):271–293

30. OTCBVS WS Series Bench, I.: http://www.cse.ohio-state.edu/otcbvs-bench/. DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968; DOD/TACOM/NAC/ARC Program under grant R01-1344-18; FAA/NSSA grant R01-1344-48/49; Office of Naval Research under grant #N000 143010022.3

31. OTCBVS WS Series Bench; Roland Miezianko, I.: Terravic research infrared database. http://www.cse.ohio-state.edu/otcbvs-bench/

32. Zhang Y, Zhou Z (2010) Cost-sensitive face recognition. IEEE Trans Pattern Anal Mach Intell 32(10):1758–1769