This version has a small correction in the abstract compared to the original version presented in IbPRIA 2013.

# Thermal Infrared Face Segmentation: A New Pose Invariant Method

Sílvio Filipe and Luís A. Alexandre

IT - Instituto de Telecomunicações
Department of Computer Science
University of Beira Interior, 6200-001 Covilhã, Portugal.
sfilipe@ubi.pt,lfbaa@di.ubi.pt

**Abstract.** This paper presents a method for automatic segmentation of images of faces captured in Long Wavelength Infrared (LWIR), allowing a wide range of face rotations, expressions and artifacts (such as glasses and hats). The paper presents a novel high accurate approach and compares its performance against three other previously published methods. The proposed approach uses active contours, morphological filtering and several other image processing operations. Experiments were performed on a total of 699 test images from three 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.

**Keywords:** Face Segmentation, Human Skin Segmentation, Image Segmentation, Infrared Thermal

## 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 light variations. In order to solve this problem, the proposed solutions are the use of 3D face recognition or a combination of facial recognition in visible and Infrared (IR) spectrum [1].

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 LWIR recognition is not affected by light variations. Contrary to the visible wavelength, where there are numerous methods for accomplishing this task (based on geometry, color, etc.) in the LWIR there is a lack of proposals to improve the current status.

A crucial step in the process of facial recognition is face segmentation. 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.

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

## 2 Overview of Face Segmentation in Thermal Infrared Images

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

Gyaourova et *al.* in [2] 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.* [3] achieve face segmentation through a Bayesian approach, fitting two Normal Distributions per class applying an adaptation of the Expectation Maximization (EM) algorithm. This algorithm accepts skin ($s$) and background ($b$) pixels from selected sub-regions of the training set where only one of those types is present, then produces 4 means ($\mu$), 4 variances ($\sigma^2$) and 2 weights ($\omega$). In the segmentation stage, for each pixel there is a prior distribution $\pi^{(t)}(\theta)$, for the $t^{th}$ iteration, where $\theta$ 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 $\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$.

The input pixel value $x_t$ has a conditional distribution $f(x_t|\theta)$. If the particular pixel is skin, we have $f(x_t|s) = \sum_{i=1}^{2} \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma^2_{s_i})$, where $\mathcal{N}(\mu_{s_i}, \sigma^2_{s_i})$ is the Normal Distribution with mean $\mu_{s_i}$ and variance $\sigma^2_{s_i}$, and where $\omega_{s_2}$ is given by $1 - \omega_{s_1}$.

More recently, Cho et *al.* in [4] 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.

Filipe and Alexandre [5] published an extension of method [3]. 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.

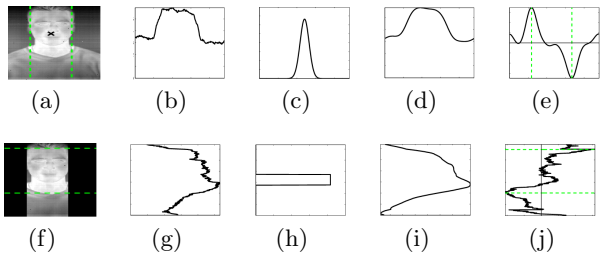


**Fig. 1.** Vertical and horizontal signatures analysis (see text for detailed description).

## 3 Proposed Method

We designed a method that is able to overcome the main problems identified with the existing face segmentation methods. The method proposed in this paper was designed to give more importance to accuracy than to speed. In the following sub-sections we will describe the steps used in our proposed method.

### 3.1 Rectangular Region of Interest (RROI)

An interesting operator would give the Rectangular Region of Interest (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 may difficult the pixel intensity-based segmentation approaches.

To obtain the RROI we 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 $sigV(c) = \sum_{r=1}^{R} I(c,r)$ and $sigH(r) = \sum_{c=1}^{C} I(c,r)$, where $c$ and $r$ are the indexes of column and row for image $I$ of dimension $C \times R$. In the first step, we start by analyzing the vertical signature (see figure 1(b)). This signal has several high frequency oscillations that will appear in its derivative (shown in figure 1(e)). This can be avoided by smoothing it with a 1D Gaussian filter (we used the one in figure 1(c) and the result shown in figure 1(d)). The following step consists in determining the extrema of this signal: in figures 1(a) and 1(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.

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 previously. 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 figure 1)) is similar to the one used to analyze the vertical signature. The main difference is the shape and size of the filter, were

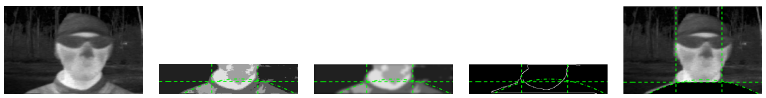


**Fig. 2.** Process for fitting a parabola (see text for detailed description).

selected to remove sudden variations that would appear in the signal when the subjects are wearing glasses or have a cold nose.

The upper bound of the face ($rowUp$) is given by the maximum of horizontal signature and is represented by a upper dash line in figure 1(f). 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 (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 in order to enhance the chin, neck and shoulders regions and remove certain types of background noise (shown in figure 2(b)). To smooth the abrupt changes in the regions presented we apply a Gaussian blur (shown in figure 2(c)). We used the Canny edge detector [6] to obtain the points used to adjust our parabola.

The result obtained by the Canny edge detector is used to fit a second order function (parabola) to obtain the parameters $a$, $b$ and $c$, with $f(x) = ax^2 + bx + c$. To find the parameters of the parabola ($a$, $b$ and $c$) which best describes the curvature of the shoulders (figures in the figure 2) or carried by the chin when the shoulders are not detected in the contours we use the least-squares method. The fit of the parabola is done automatically where it receives a set of points, in this case the points are the edges returned by the Canny method (white points shown in the figure 2(d)). $rowDown$ is given by the mean value between $f(colLeft)$ and $f(colRight)$ when $a > 0$ (adjusts to the chin) and when $a < 0$ $rowDown$ is given by the maximum value of the parabola (adjusts to the shoulders).

### 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 in [2] 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 section 3.3). 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 figure 1(a), we use the extrema obtained while searching for the RROI.

(a) (b) (c) (d) (e) (f)

**Fig. 3.** Step by step example of the application of the FPIBI operation (see text for detailed description).

### 3.3 Active Contours without Edges

Chan and Vese [7] 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. Since it is an iterative method, 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. The processing time will depend on the type of initial contour and of its position. Our method will use an active contour but with an elliptical initial boundary, thus having a shape similar to a face.

### 3.4 Face Pixel Identification from Binary Image (FPIBI)

To select the face pixels in binary images (see figure 3(a)) that result from the application of the method [7]. 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 figure 3). Firstly we identify the center of the face as explained in the RROI operation (cross in the figure 3(a)), followed by a morphological operation used to remove small areas. Then, an edge map is obtained using the Canny edge detector [6] and enhanced through a dilation (see figure 3(b)). From these edges we select and fill the largest that contains the face center (see figure 3(c)).

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 figure 3(d)). Then we apply an opening with a circular structuring element (see figure 3(e)). Only the largest regions, such as the glasses, remain after the application of this morphological operator. This resulting image is summed with the one that results from filling the largest contour and the values that exceed the largest possible value are clipped to zero (non-face pixel) (see figure 3(f)).

## 4 Experimental Results

### 4.1 Datasets

In these work, we used three database the face images having different characteristics in the LWIR and these are: 1 – The 'Collection X1' University of Notre Dame (UND) database [8] contains 2293 LWIR frontal face IR images from 81 different subjects. The training set contains 159 images and the test set 163. 2 – The 'Terravic Facial IR Database' [9] contains 24508 images of 20 different persons. It has different poses and artifacts, and the images was captured in an indoor and outdoor environment, for this the training set has 235 images and the test set has 240. 3 – The 'IRIS Thermal/Visible Face Database' [10] contains 4228 images were acquired with various rotations, illuminations and expressions and in this database the training and test sets have both 296 images.

The test sets images from all databases were segmented manually to create the test set ground truth (samples shown in row 2 of figure 4). Method [4] does not need a training set and method [3] and ours use pixels from manually segmented regions of the training set images avoiding the need for accurate segmentation of the training set. A list with the names of the images used for training, test, the segmented images and all the source code is available at: `http://socia-lab.di.ubi.pt/~silvio/IbPRIA2013.html`.

### 4.2 Results and Discussion

The test set of the databases were used to measure pixel-by-pixel agreement between the binary maps produced by each of the algorithms (these maps are shown in figure 4, rows 3 to 5) and the ground-truth data manually built *a priori* (see examples in row 2 of figure 4).

During this work, we evaluate the accuracy of each algorithm by measuring the errors $E_1$ and $E_2$. The quantitative evaluation of the presented methods is in table 1. It contains the errors $E_1$ and $E_2$, of each algorithm and the execution time.

The classification error rate ($E_1$) of the algorithm on the input image $I_i$ is given by the proportion of correspondent disagreeing pixels across the image: $E_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{C\times R}\sum_{c=1}^{C}\sum_{r=1}^{R} O(c,r) \otimes T(c,r)$, where $O(c,r)$ and $T(c,r)$ are, respectively, pixels of the output and true class images.

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 rate of the images is given by the average between the False Positive Rate (FPR) and False Negative Rate (FNR): $E_2 = \frac{1}{n}\sum_{i=1}^{n}\frac{FNR}{2} + \frac{FPR}{2}$.

Regarding the results for the UND database (in table 1), our method improves them between 4.4% and 31.3%. Method [3] only analyzes the distribution of intensities of the pixels, so that when there is a region with clothes that has a temperature similar to the skin, it is considered to be skin. The second part of table 1 shows the results of the methods for segmenting the Terravic database

**Table 1.** Experiment results in all databases. For each method, we present the $E_1$, $E_2$ and execution time in seconds. The best (smallest) results in terms of error measures for each database are shown in bold.

| | | Methods | | |
|---|---|---|---|---|
| | | Cho et *al.* [4] | Pavlidis et *al.* [3] | Proposed Method |
| UND | $E_1$ | 0.369 | 0.118 | **0.074** |
| | $E_2$ | 0.376 | 0.143 | **0.063** |
| | $Time(s)$ | 1.236 | 4.480 | 5.974 |
| Terravic | $E_1$ | 0.158 | 0.095 | **0.047** |
| | $E_2$ | 0.179 | 0.083 | **0.057** |
| | $Time(s)$ | 1.237 | 4.554 | 7.041 |
| IRIS | $E_1$ | 0.307 | 0.181 | **0.096** |
| | $E_2$ | 0.451 | 0.156 | **0.104** |
| | $Time(s)$ | 0.959 | 3.233 | 7.686 |

and the improvement of our method range from 2.6% to 12.2% (error measure $E_2$). Our method 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 table 1. All methods had an increase in error rates for both error measures ($E_1$ and $E_2$). In this database, the method in [3] 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. In our method, the increase was not as sharp since the cuts were made solely by the fit of the parabola. Still, our method improved the segmentation results from 5.2% to 34.7% (in measure $E_2$).

The third row of figure 4 contains the result of the method in [4]. We observe that when the extracted contour is not closed, it assumes that much of the background is part of the face. In the fourth row of figure 4 is shown the segmentation result of the method in [3]. This method is based on a model of the distribution of intensities of the pixels, which means that all regions that have a higher temperature (higher intensity of pixels) are assumed as part of the face. The results of our method can be seen in row 5 of figure 4. In order to try to solve the problems presented by the other methods, we defined steps strategically targeted to these problems. Through the combination of several operations that are included in this method it can approximate quite well the desired results of manually segmented images.

## 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 a new method that performs well regardless of face pose, rotation, artifacts

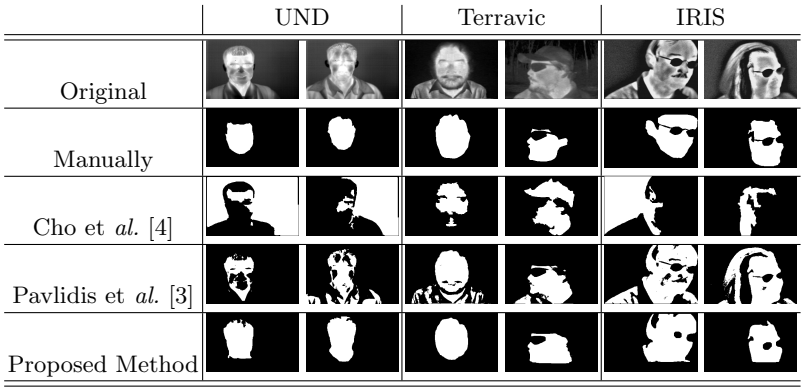


**Fig. 4.** Sample test set results for images from all databases.

and expression, and an extensive evaluation of these methods under different publicly available databases (690 training and 699 test images).

The proposed method was designed to be accurate. It does this quite well since it is the best in both error measures, with improvements of up to 29.5% according to $E_1$ and 34.7% according to $E_2$, depending on the database. In 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.

## Acknowledgments

'FCT - Fundação para a Ciência e Tecnologia' and 'FEDER' in the scope of the PTDC/EIA-EIA/103945/2008 research project 'NECOVID: Covert Negative Biometric Identification' and in the PTDC/EIA/69106/2006 research project 'BIOREC: Non-Cooperative Biometric Recognition' are acknowledged.

## References

1. Kong, S., Heo, J., Abidi, B., Paik, J., Abidi, M.: Recent advances in visual and infrared face recognition - a review. Computer Vision and Image Understanding **97**(1) (2005) 103–135
2. Gyaourova, A., Bebis, G., Pavlidis, I.: Fusion of infrared and visible images for face recognition. In: European Conference on Computer Vision. (2004) 456–468
3. Pavlidis, I., Tsiamyrtzis, P., Manohar, C., Buddharaju, P.: Biometrics: Face recognition in thermal infrared. In: Biomedical Engineering Handbook. 3 edn. CRC Press (2006) 1–15
4. Cho, S., Wang, L., Ong, W.: Thermal imprint feature analysis for face recognition. IEEE International Symposium on Industrial Electronics (2009) 1875–1880

5. Filipe, S., Alexandre, L.A.: Improving Face Segmentation in Thermograms Using Image Signatures. In: 15th Iberoamerican Congress on Pattern Recognition, São Paulo, Brazil, Springer (2010) 402–409
6. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence **8**(6) (September 1986) 628–633
7. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Transactions on Image Processing **10**(2) (2001) 266–277
8. Chen, X., Flynn, P., Bowyer, K.: IR and visible light face recognition. Computer Vision and Image Understanding **99** (2005) 332–358
9. Miezianko, R.: Terravic Research Infrared Database. (2006) http://www.cse.ohio-state.edu/otcbvs-bench/.
10. Abidi, B.: IRIS Thermal/Visible Face Database. (2007) DOE University Research Program in Robotics under grant DOE-DE-FG02-86NE37968.