

# Improving Face Segmentation in Thermograms Using Image Signatures

Silvio Filipe and Luís A. Alexandre

Department of Computer Science  
IT - Instituto de Telecomunicações  
SOCIA - Soft Computing and Image Analysis Group  
University of Beira Interior, 6200-Covilhã, Portugal  
{m2213,lfbaa}@ubi.pt

**Abstract.** The aim of this paper is to present a method for the automatic segmentation of face images captured in Long Wavelength Infrared (LWIR), allowing for a large range of face rotations and expressions. The motivation behind this effort is to enable better performance of face recognition methods in the thermal Infrared (IR) images. The proposed method consists on the modelling of background and face pixels by two normal distributions each, followed by a post-processing step of face dilation for closing holes and delimitation based on vertical and horizontal images signatures. Our experiments were performed on images of the University of Notre Dame (UND) and Florida State University (FSU) databases. The obtained results improve on previous existing methods from 2.8% to more than 25% depending on the method and database.

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

## 1 Introduction

A large amount of research has been conducted in the field of face recognition, mainly in the visible spectrum. These systems have problems dealing with light variations [6]. Some of the proposed solutions use 3D facial recognition [1] and combine face recognition in both visible and IR spectrum [7].

The growing interest in robust methods (for example, for security applications) has driven the development of facial recognition exclusively in the infrared. Recognition in the LWIR is not affected by light variations.

A crucial step in the process of face recognition is the face segmentation. This is more demanding than simple face detection since it pinpoints not only the face's locations, but also must describe its shape. A robust segmentation system can improve recognition rates regardless of the recognition method.

In contrast with the visible wavelength, where numerous methods have been proposed to accomplish this task (based on color, geometry, etc.), in the LWIR there is a lack of proposals to improve the current status.

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

## 2 Overview of Face Segmentation in Thermal Infrared Images

Face segmentation, given that it is a preprocessing step for all recognition methods, will lead to their failure if it is not correctly performed. This is not a subject much discussed by the authors of recognition methods in the infrared. Some of the proposed approaches are based only on the creation of an elliptical mask that will be put over the image of the face [5], but these approaches will work only on frontal and centered faces.

*Siu-Yeung Cho et al.* in [3] present a method for face segmentation in IR images based on the Sobel Edge detector and morphological operations. After the Sobel Edge detector, the largest contour is considered to be the one best describing the face. They apply the morphological operations to the area contained in this outline to connect open contours and remove small areas. Figures 2(e) and 2(f) show the segmented images in figures 2(a) and 2(b) using this method.

*I. Pavlidis et al.* in [8] describe a method for face segmentation using a Bayesian Approach. This method is based on the combination of two Normal Distributions per class, which are estimated using the Expectation-Maximization (EM) algorithm. This algorithm uses pixels from the skin ( $s$ ) and background ( $b$ ) for training. These are obtained from the training set images by selecting subregions that contain only pixels from each of these types. With this, the EM returns 4 means ( $\mu$ ), 4 variances ( $\sigma^2$ ) and 4 weights ( $\omega$ ).

In the segmentation stage, for each pixel they have a prior distribution ( $\pi^{(t)}(\theta)$  where  $t$  is the iteration) to whether that pixel is skin ( $\pi^{(t)}(s)$ ) or background ( $\pi^{(t)}(b) = 1 - \pi^{(t)}(s)$ ).  $\theta$  is the parameter of interest, which takes two possible values ( $s$  and  $b$ ) with some initial (prior) probability ( $\pi^{(1)}(s) = \frac{1}{2} = \pi^{(1)}(b)$ ).

The input pixel value  $x_t$  has a conditional distribution  $f(x_t|\theta)$  and if the particular pixel is skin we have:

$$f(x_t|s) = \sum_{i=1}^2 \omega_{s_i} \mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2) \quad (1)$$

where  $\mathcal{N}(\mu_{s_i}, \sigma_{s_i}^2)$  is the Normal Distribution with mean  $\mu_{s_i}$  and variance  $\sigma_{s_i}^2$ .

The prior distribution ( $\pi^{(t)}(\theta)$ ) combined with the likelihood ( $f(x_t|\theta)$ ) provides (via the Bayes theorem) the posterior distribution ( $p^{(t)}(\theta|x_t)$ ), where, for the skin pixels, according to the Bayes theorem, we have:

$$p^{(t)}(s|x_t) = \frac{\pi^{(t)}(s)f(x_t|s)}{\pi^{(t)}(s)f(x_t|s) + \pi^{(t)}(b)f(x_t|b)} \quad (2)$$

The posterior distribution is also used to obtain the prior distribution for the next iteration:

$$\pi^{(t+1)}(\theta) = \begin{cases} \pi^{(t+1)}(s) = p^{(t)}(s|x_t), & \text{when } \theta = s \\ \pi^{(t+1)}(b) = 1 - \pi^{(t+1)}(s), & \text{when } \theta = b \end{cases} \quad (3)$$

Figures 2(g) and 2(h) show the segmented images in figures 2(a) and 2(b) using this method.

### 3 Proposed Method

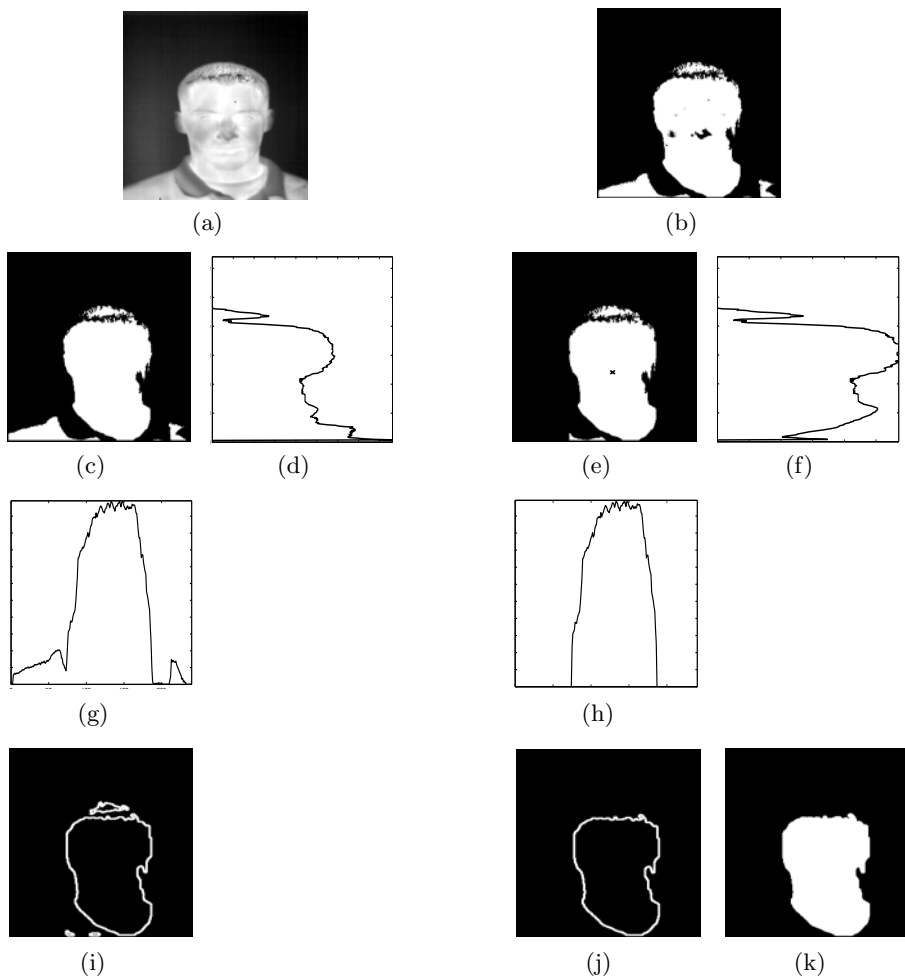
We evaluated the methods of [3] and [8] and realized that it was possible to improve their results. Our proposal is based on the method of [8]: after analyzing the results of this method (shown in figure 1(b)) we concluded that its main problem is the removal clothing because since the body warms it, clothes have temperatures similar to the skin.

To remove the effect of the clothing, we examined the vertical and horizontal pixel signatures (see figure 1(d) and 1(g)). The vertical and horizontal signatures are the sum of the pixels values along the columns, for the vertical signature, and columns, for the horizontal signature. After this we fill small areas (shown in figure 1(c)) using a dilation with a  $4 \times 4$  filter. This enables the removal of incorrectly classified pixels using a proportion of the maximum values in the two signatures (see figure 1(e)). This proportion is 20%, i.e., all signatures that have values below 20% of the maximum value are considered as background. This value was obtained searching for the best performance in the training sets of the databases.

After this, we calculate the possible location of the center of the face with new signatures (horizontal in figure 1(f) and vertical in 1(h)) in figure 1(e). The center point is given by the maximum values of the signatures (when more than one maximum value exist in the horizontal or vertical signatures the average of these maximums is used). This possible center location of the face (marked with a cross in figure 1(e)) will be used for the search for the largest contour (see figure 1(j)). Before we look for the largest contour, we apply an erosion followed by dilation with a filter of  $3 \times 3$  and  $2 \times 2$ , respectively. This is used to remove some links between areas. For contour extraction we use the Canny edge detector. To enhance the contours extracted by the Canny method used a dilation with a  $3 \times 3$  filter (see figure 1(i)). Only boundaries that have the center point inside are accepted, producing the end result of figure 1(k).

A possible drawback of this method occurs when the calculated center position of the face is not correct. This may cause the largest contour to be only partially over the actual face.

Figures 2(i) and 2(j) show the segmented images in figures 2(a) and 2(b) using our method.



**Fig. 1.** Thermal face segmentation process. Figure 1(a) is the original image, from the training set of the UND database. Figure 1(b) is the original image (figure 1(a)) segmented by the method [8]. Figure 1(c) is the image (figure 1(b)) after filling small areas. Figure 1(d) is the horizontal signature of figure 1(c). Figure 1(e) is the result of the analysis of horizontal and vertical signatures of figure 1(c). Figure 1(f) is the horizontal signature of figure 1(e). Figure 1(g) is the vertical signature of figure 1(c). Figure 1(h) is the vertical signature of figure 1(e). Figure 1(i) is the result of enhancement (using a dilation with a  $3 \times 3$  filter) of the contours extracted from the figure 1(e) by the Canny edge detector. Figure 1(j) is the largest contour of the figure 1(i). Figure 1(k) is the result of the face segmentation in the original image (figure 1(a)) using our method, after filling the area inside the contour of the figure 1(j).

## 4 Experimental Results

### 4.1 Datasets

The UND database is presented in [4,2]. The ‘‘Collection C’’ of the UND database contains 2293 LWIR frontal face IR images from 81 different subjects. The training set contains 159 images and the test set 163.

The FSU database contains 234 frontal IR images of 10 different subjects, which were obtained at varying angles and facial expressions [9]. The train set contains 40 IR images (four per subject) and the test set 194. The images from this database have color representation. The color channels (Red (R), Green (G) and Blue (B)) and grayscale conversion were processed separately, i.e., the algorithm process R, G, B and grayscale independently.

All test set images from both databases were segmented manually to create the ground truth for test sets. Method [3] does not need a training set and method [8] and ours use pixels from manually segmented regions of the training set images avoiding the need for accurate segmentation of the training set.

### 4.2 Experimental Results and Discussion

The requested task is quite simple: for each input image (see figure 2(a) and 2(b)) a corresponding binary output (shown in figure 2(c) and 2(d)) 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 were used to measure pixel-by-pixel agreement between the binary maps produced by each of the algorithm  $O = O_1, \dots, O_n$  (images in figures 2(e), 2(f), 2(g), 2(h), 2(i) and 2(j)) presented earlier and the ground-truth data  $C = C_1, \dots, C_n$ , manually built *a priori* (shown in figure 2(c) and 2(d)).

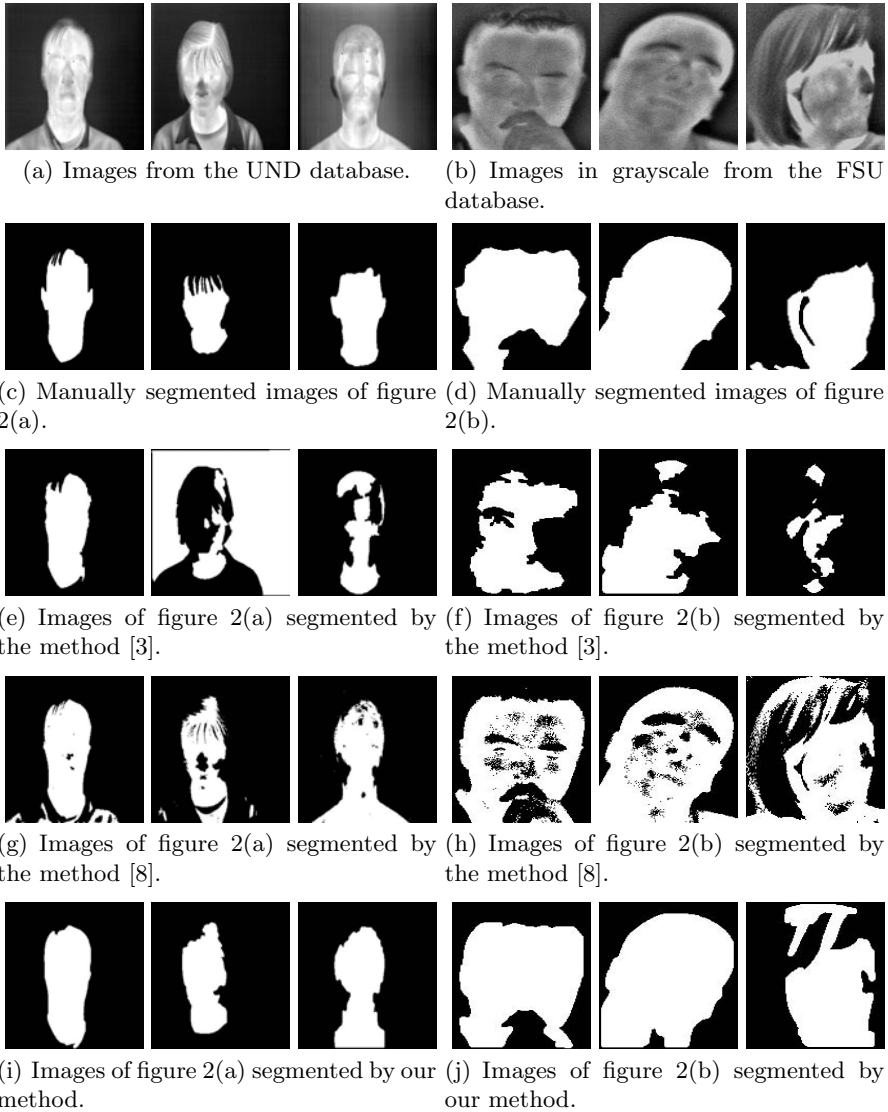
The classification error rate ( $E^1$ ) of the algorithm is given by the proportion of correspondent disagreeing pixels (through the logical exclusive-or operator, see equation 4) over all the image, where  $O(c', r')$  and  $C(c', r')$  are, respectively, pixels of the output and true class images.

$$E^1 = \frac{1}{c \times r} \sum_{c'} \sum_{r'} O(c', r') \otimes C(c', r') \quad (4)$$

The second error measure 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 ( $E^2$ ) of the images is given by the average between the False Positive Rate (FPR) and False Negative Rate (FNR).

$$E^2 = 0.5 \times FNR + 0.5 \times FPR \quad (5)$$

The results of segmentation for the described methods are presented in table 1. For the UND database, we can observe that error rates obtained with our method improved upon the results of the other two methods. The same was not the case for the FSU database. In this, the FPR increased, but the final



**Fig. 2.** Input images for the two databases, manually segmented images and the segmented images by the three methods

two errors ( $E^1$  and  $E^2$ ) decreased due to the FNR decrease significant. In this database the FPR's increased because when the subject wears glasses and we dilate the images from the method described in [8] we include part of the glasses as face pixels and they do not belong to the face. For the FSU images (shown in figure 2(b)), most of the noise comes from the hair since that the face fills almost entirely the image unlike what happens in the UND database, where there is a large area not covered by the face (see figure 2(a)).

**Table 1.** Segmentation methods results in the test sets from the UND and FSU databases. The FSU  $Fusion^1$  is the fusion between the results of segmentation for the R, G and B channels using the majority vote between them. The FSU  $Fusion^2$  is the fusion of the results of grayscale and the R, G and B channels using a weight of 0.3 for the grayscale channel and  $\frac{0.7}{3}$  for the other.

	<i>Siu-Yeung Cho et al.[3]</i>				<i>I. Pavlidis et al.[8]</i>				<i>Ours</i>			
	FNR	FPR	$E^1$	$E^2$	FNR	FPR	$E^1$	$E^2$	FNR	FPR	$E^1$	$E^2$
UND	0.369	0.354	0.356	0.362	0.166	0.080	0.093	0.123	0.145	0.050	<b>0.065</b>	<b>0.097</b>
FSU Gray	0.502	0.048	0.308	0.275	0.200	0.174	0.189	0.187	0.044	0.206	<b>0.114</b>	<b>0.125</b>
FSU R	0.533	0.046	0.325	0.290	0.206	0.169	0.190	0.187	0.058	0.186	<b>0.112</b>	<b>0.122</b>
FSU G	0.502	0.047	0.307	0.275	0.171	0.184	0.171	0.178	0.044	0.206	<b>0.114</b>	<b>0.125</b>
FSU B	0.366	0.085	0.246	0.226	0.170	0.186	0.177	0.178	0.033	0.242	<b>0.123</b>	<b>0.138</b>
FSU $Fusion^1$	0.494	0.048	0.303	0.271	0.171	0.184	0.177	0.178	0.045	0.206	<b>0.114</b>	<b>0.125</b>
FSU $Fusion^2$	0.482	0.049	0.296	0.265	0.171	0.184	0.177	0.178	0.045	0.206	<b>0.114</b>	<b>0.125</b>

The database FSU was analyzed for each channel independently and two fusions ( $Fusion^1$  and  $Fusion^2$ ) were made to verify what would be the best approach for the segmentation.

The FSU  $Fusion^1$  is the fusion between the results of segmentation for the R, G and B channels using the majority vote between them. The FSU  $Fusion^2$  is the fusion of the results of grayscale and the R, G and B channels using a weight of 0.3 for the grayscale channel and  $\frac{0.7}{3}$  for the other.

For the method [3] the best result was obtained with the blue channel (FSU B) for both errors  $E^1$  and  $E^2$ . For the method [8] the smallest error  $E^1$  is obtained with the green (FSU G) and the smallest error  $E^2$  appears in the green (FSU G) and blue (FSU B) channels and for the FSU  $Fusion^1$  and FSU  $Fusion^2$ .

Our method has the best result in the red channel (FSU R) for both errors  $E^1$  and  $E^2$ . With this, we can say the best result of each method in this database depends on the type of images and that fusions does not always improve the results.

The improvements brought by our approach in relation to method [8] are the removal of clothing and the inclusion of larger number of pixels of the face. Removal of clothing is quite visible in the difference between the images resulting from [8] (figure 2(g)) and our method (figure 2(i)). In these examples it is possible to see that almost all the clothes were removed in the images of the UND database. The inclusion of the pixels of the face is most visible in the FSU database as can be seen in the images of figures 2(h) and 2(j), the first being the result of [8] and ours is the second. With this we minimize the FNRs, causing us to obtain more pixels for face recognition tasks.

## 5 Conclusion

In this paper we proposed a face segmentation method for LWIR images. The method creates two Gaussian distributions for each type of pixel (face and background) and post-processes the obtained images by closing small holes using morphological operators (decreasing FNRs) and removing the effect of clothes through the analysis of vertical and horizontal image signatures.

The experimental results show that our proposal improves accuracy from 2.8% to over 25% depending on the dataset and the method against which we are comparing.

We are currently searching for new features that may improve segmentation performance.

## Acknowledgments

We acknowledge the financial support given by “FCT - Fundação para a Ciência e Tecnologia” and “FEDER” in the scope of the PTDC/EIA/69106/2006 research project “BIOREC: Non-Cooperative Biometric Recognition”.

We wish to thank Professor Cho Siu-Yeung David, Assistant Professor in the School of Computer Engineering at Nanyang Technological University (NTU) for the source code of his method[3].

## References

1. Bowyer, K., Chang, K., Flynn, P.: A survey of approaches to three-dimensional face recognition. In: 17th International Conference on Pattern Recognition (ICPR 2004), pp. 358–361 (2004)
2. Chen, X., Flynn, P., Bowyer, K.: IR and visible light face recognition. *Computer Vision and Image Understanding* 99, 332–358 (2005)
3. Cho, S., Wang, L., Ong, W.: Thermal imprint feature analysis for face recognition. In: IEEE International Symposium on Industrial Electronics (ISIE), pp. 1875–1880
4. Flynn, P., Bowyer, K., Phillips, P.: Assessment of Time Dependency in Face Recognition: An Initial Study. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 44–51. Springer, Heidelberg (2003)
5. Gyaourova, A., Bebis, G., Pavlidis, I.: Fusion of infrared and visible images for face recognition. In: Pajdla, T., Matas, J. (eds.) ECCV 2004, Part IV. LNCS, vol. 3024, pp. 456–468. Springer, Heidelberg (2004)
6. Jain, A., Flynn, P., Ross, A.: *Handbook of Biometrics*. Springer, New York (2007)
7. 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* (1), 103–135
8. Pavlidis, I., Tsiamyrtzis, P., Manohar, C., Buddhharaju, P.: *Biometrics: Face recognition in thermal infrared*. ch. 29, pp. 1–15. CRC Press, Boca Raton (2006)
9. Srivastava, A., Liu, X.: Statistical hypothesis pruning for identifying faces from infrared images. *Image and Vision Computing*, 651–661