

ROIs Segmentation in Facial Images Based on Morphology and Density Concepts

Jesús García-Ramírez^(✉), J. Arturo Olvera-López, Ivan Olmos-Pineda,
and Manuel Martín-Ortíz

Faculty of Computer Science, Benemérita Universidad Autónoma de Puebla,
Puebla, Mexico

gr_jesus@outlook.com, {aolvera,iolmos,mmartin}@cs.buap.mx

Abstract. In computer vision, facial images have several applications such as Facial Expression Recognition and Face Recognition. The segmentation of Regions Of Interest (ROIs) in face images are relevant, because those provide information about facial expressions. In this paper a method to segment mouth and eyebrows in face images based on edge detection and pixel density is proposed. According to the experimental results, our approach extracts the ROIs in face images taken from different public datasets.

Keywords: Face images · Face ROIs segmentation · Image processing · Expression recognition

1 Introduction

Nowdays, computer vision is an interesting area for computing researchers, where image processing is the baseline for that area. Some examples of applications in computer vision are: facial expression recognition and face recognition. Different works related to these applications have been reported in the literature [1, 2].

The human face is a part of the body with a great scientific interest, because of many expression such as angry, happiness, fear, among others, are reflected on this region. Regions Of Interest (ROIs) in the human face are the eyes, nose, eyebrows, and mouth, these regions describe features from human expressions. Finding these ROIs in digital images is not an easy task because of the low contrast between skin color and those ROIs.

For the last reason, the image processing filters such as thresholding and border extraction do not work well at all in face images, those filters need to be improved in order to get a better performance in this kind of images.

Border extraction in ROIs is a very important task in order to get descriptive data from the face; this information can be used for either face recognition or to find facial expressions. It is difficult to apply a global analysis to face images, because of the ROIs have different features about illumination and density, then a regional analysis is a better way for detecting borders in ROIs for face images.

For mouth detection, different approaches have been proposed, for example, chromatic information and the Expectation-Maximization algorithm are used to segment the mouth in face images [3]. A different method analyzes image histogram, where ROIs are detected from color energy [4], using R and G channels from the RGB color model. Other approaches use information such as blood concentration [5], where regions such as lips are detected. The disadvantage of this strategy is the fact that it is necessary to capture the input images with a monochrome camera equipped via an acousto-optic device that captures blood concentration. A different strategy for detecting lips and mouth is based on active shape models and active contour models [6], which is robust to different conditions of illumination.

On the other hand, for detecting eyebrows, different authors propose a binarization strategy, using different color space brands (L and b from CIELAB color space), then the Otsu algorithm is applied to find the eyebrow region [7]. In [8] a method based on local active shape model is proposed, in this approach different angles and distances between the eyes and the eyebrows are used to find the shape of a ROI.

In this paper a method for segmenting mouth and eyebrows is proposed, where the aim is to minimize the error (noise pixels detected as ROI, and ROI pixels detected as noise). It is important to mention that eyebrows and mouth contribute with relevant information about facial expressions. The method consist of applying different pixel operations, such as edge detectors and contrast modification, in order to obtain a binary image. Finally, a method based on density and morphological techniques is proposed for getting descriptive information from ROIs including a stage for eliminating noise regions.

This paper is structured as follows: in Sect. 2 the proposed methodology is presented; in Sect. 3 the experimental results obtained by applying our approach over three public datasets are introduced. Finally, conclusions and future work are discussed in Sect. 4.

2 Methodology for Face ROIs Segmentation

In this section a couple of algorithms to find the ROIs in facial images (mouth and eyebrows) are introduced. The proposed methodology can be seen in Fig. 1 and it is described in the following paragraphs. We assume that input digital images are represented according to the RGB color model. However, in our process, channels R and G for the gray scale conversion are used, because B channel does not provide information about edges [3]. The face is located by the Viola & Jones algorithm, the eyebrows and mouth regions are segmented, finally the obtained regions are denoised. Details about our proposed algorithms are described in the following lines.

Digital image processing is a computational expensive process, because all the pixels in the image are taken into account to apply a filter, this process depends on the image size. For that reason, segmentation is an important task to process fewer pixels in the image. In order to reduce the ammount of pixels

to process, the face region is located with the Viola & Jones algorithm which is based on intensities of the pixels related to mouth and eyes regions, which are darker than another around them [9].

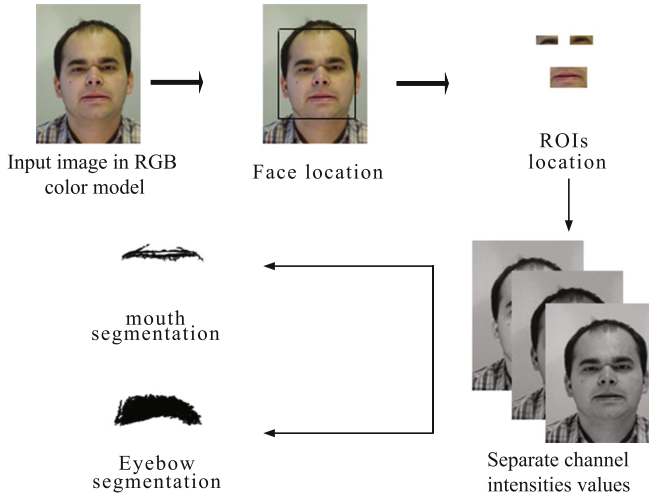


Fig. 1. Proposed Methodology which consists of the stages: Face location, ROIs Location, ROIs segmentation. Input images were taken from the MMI database [12].

In order to find mouth and eyebrows regions in the image, a template with values related to initial and final points of a rectangular area are proposed and they are shown in Table 1, where width and height are values from the output image by the Viola & Jones algorithm. Before getting the values of the Table 1, other values were tested, the better values are presented in the Table 1, these values are proposed for the used databases according to the geometrical features; these values can be modified depending on the dataset.

After the template is used to locate the ROIs, the next step is to apply edge detectors and filters in order to find edge information for mouth and eyebrow regions. These regions have different features, the mouth is a region where edge detectors are applied to get descriptive information about that, on the other hand eyebrow region is not a dense region because shadows between it and eye region are visible in a face image; due these facts different approaches to get descriptive information from each region need to be applied.

In face images the transitions between the ROIs such as mouth and the skin color are not visible, for this reason, the regional filters need to be modified, to take into account a higher area for applying the convolution process, in order to have a better performance in edge detection.

For detecting edge information in face images, we propose an extended convolution matrix, considering dimensions $(2u + 1)$ by $(2u + 1)$, where u is the value that determine the matrix dimension. In Fig. 2 it is shown an example

Table 1. Points to process the template to find the eyebrow and mouth regions in face images.

	Initial point of the region	Final point of the region
Mouth	$(height * .6, width * .2)$	$(height, height * .8)$
Rigth eyebrow	$(height * .2 + width * .1)$	$(height * .4, height * .5)$
Left eyebrow	$(height * .2 + width * .5)$	$(height * .4, height * .9)$

with $u = 1$ and $u = 2$ related to the Sobel convolution matrix (which is used in our experiments).

In our approach for mouth segmentation, intensity values of both green and blue channels are processed. The EDEM (Edge DETection in Mouth) algorithm (see Algorithm 1) has as input a RGB image, the first step separates the three channels, for the mouth region only R and G channels are processed. To find the edges in mouth regions the gradient is applied using the convolutional masks (horizontal and vertical directions) shown in expression 1 and considering $u = 2$ according to Fig. 2. After obtaining the edges, they are enhanced by applying the sine filter. The next step is to analyze the histogram taking into account the intensities that have a higher frequency value than the mean of the histogram (lines 8–15), after this process, the intensity values are mostly located either near to zero (low regions) or near to 255 (high intensities), then a binarization process ($x = 0$ if $x \leq threshold$, $x = 255$ otherwise) is applied with a $threshold = 127$, finally the algorithm returns an image with the edges of the mouth and some noise, this noise will be removed in the next step.

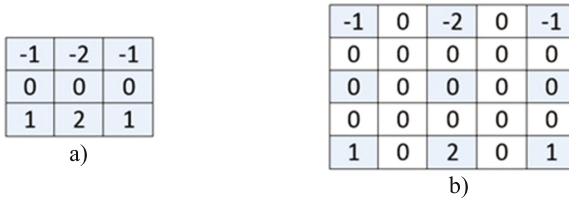


Fig. 2. Modification of the convolution matrices used in our approach, (a) Region with $u = 1$, (b) Region with $u = 2$.

$$\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{1}$$

Now the eyebrows region will be segmented, for this region other process is applied because the features like the density of the region or the lightning are different among them. In this process only the R channel is used to find the region.

Algorithm 1. Edge DETection in Mouth (EDEM) algorithm.

Require: Image in RGB color model I
Ensure: Binary image with the mouth edges I_s

- 1: Separate the intensities values of channels $I=I_R, I_G, I_B$
- 2: **for** each value in I_R and I_B **do**
- 3: G_V =Apply the vertical gradient (Equation 1)
- 4: G_H =Apply the horizontal gradient (Equation 1)
- 5: value= $\sqrt{(G_V)^2 + (G_H)^2}$
- 6: The new pixel value will be: $255 * \text{Sin}(\frac{\pi * \text{value}}{2 * 255})$
- 7: **end for**
- 8: Compute the histogram (H_R, H_G) of each new images
- 9: Compute the mean of I related to the intensities values $\mu: \frac{\text{width} * \text{height}}{255}$
- 10: **for** each value in I_R and I_B **do**
- 11: **if** The frequency of the histogram value is minor than μ **then**
- 12: The new value of the pixel in I_s will be 255
- 13: **end if**
- 14: **end for**
- 15: Binarization with $\text{threshold} = 127$
- 16: Return I_s

For segmenting the eyebrows, ERED (Eyebrow REgion Detection) algorithm is proposed and it is shown in Algorithm 2. This algorithm takes as input a RGB image but only the R channel is processed, to increase the contrast the hyperbolic tangent filter is applied followed by the thresholding function in Eq. 2 where f' is the sine filter, this process is applied to remove the shadow between the eye and the eyebrow region.

$$f(x, y) = \begin{cases} f(x, y) & \text{if } f(x, y) < \text{threshold} \\ f'(x, y) & \text{if } f(x, y) > \text{threshold} \end{cases} \quad (2)$$

The eyebrow is not a dense region, for that reason a morphological closing operation is applied to smooth the contour and eliminate thin holes in the image. The structure element considered in our approach can be seen in Eq. 3, this is an element commonly used in morphological operations. Then Otsu algorithm is applied to binarize the image in order to find eyebrow information [10]. Finally, the image with some noise and the eyebrow region is returned.

$$\text{Structure Element} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad (3)$$

EDEM and ERED algorithms have as output a binary image with the ROI and some noise of other parts in the face. To denoise the image an algorithm based on the clustering algorithm DBscan [11] is proposed. This process can be seen in the Algorithm 3 (Dense Regions in Binary Image, DERBI) which has as input a binary image, in this case the black pixels are the edges information of the ROIs, the main objective of this algorithm is to minimize the noise and obtain the ROIs in a binary image. This algorithm has as output a list with the clusters in the image. The algorithm analyzes the black pixels and their

Algorithm 2. Eyebrow REgion Detection (ERED) algorithm.

Require: Image in RGB color model I
Ensure: Binary image with the eyebrow region I_s

- 1: Separate the channel of I $\{I_R, I_G, I_B\}$
- 2: Apply the hyperbolic tangent filter to I_R .
- 3: **for** each value in I_R **do**
- 4: Apply the threshold filter of the Equation 2
- 5: **end for**
- 6: Apply the closing morphological filter with the structure element of equation 3 I_M
- 7: Compute the Histogram H of I_M and apply the Otsu algorithm to get a threshold th
- 8: **for** each value x in I_R **do**
- 9: **if** x is minor than th **then**
- 10: the new pixel value will be 255
- 11: **else**
- 12: the new pixel value will be 0
- 13: **end if**
- 14: **end for**
- 15: Return the new image I_s

neighbours, taking as density those pixels surrounding with the same color (in this case black), all reachable pixels (with a distance equal to 1) are added to a list and then all the pixels in the list are analyzed in a similar way. If there is not more black pixels a new list is created and the process is repeated with other black pixel. The process finishes when all the black pixels are analyzed.

As output of DERBI algorithm a list of clusters with the coordinates of black pixels is returned, those clusters contain the dense regions in a binary image. After the list of clusters is found the next step is to apply some metrics to determine the clusters corresponding to the ROI. The used metrics are the density with respect to the rectangular area of the cluster, the region in the image related to this rectangular area and the proximity to the center of the image. A range to each metric is established to get the ROI information, if the cluster fulfils these ranges it will be depicted in the binary image.

Algorithm 3. DENSE Regions in Binary Image (DERBI) algorithm.

Require: Binary image I
Ensure: List with the clusters in I ($Clusters$)

- 1: A list with the points coordinates of the black pixels in I ($Points$)
- 2: **for** each point x in $Points$ **do**
- 3: **if** x is not visited **then**
- 4: Create a cluster and add to $Clusters$
- 5: Add the coordinates to the current cluster with their neighbours in $Points$
- 6: Set the pixels coordinates as visited
- 7: **for** each value in the current cluster **do**
- 8: Add the neighbours coordinates to the current cluster of the not visited neighbours and then set as visited
- 9: **end for**
- 10: **end if**
- 11: **end for**
- 12: Return $Clusters$

3 Experimental Results

In this section the results of applying the methods described in Sect. 2 are reported. In the experiments the MMI database was used, this database consist of 474 images, the images of the database are captured from five subjects [12]. In addition Jaffe and VidTIMIT databases are used to compare the result of our approach to segment mouth region. Jaffe Database contains 213 images from 60 Japanese subjects of 7 expressions [13], VidTIMIT contains video and audio recording from 43 persons [14].

To determine the accuracy of the approaches, a test set of the database was selected, the set consist in five images of each subject, the ROIs in the images was manually segmented to compare the segmented images with our approach. To compare the images a polynomial is found with the divided differences method, then the polynomial coefficients from the control images and the segmented images are compared. For mouth region two quadratic polynomials are computed (one for the upper region of the mouth and other for the lower one), and for the eyebrow region only one cubic polynomial is computed. An example for these two polynomials is shown in Fig. 3, the mathematical expression to get points is presented in the top of the images.

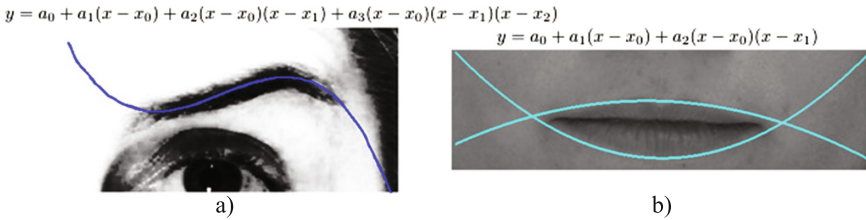


Fig. 3. Examples of interpolated functions using polynomial expressions: (a) eyebrow region and (b) mouth region.

The points of each ROI are translated to the origin, so the first point is located in $(0, 0)$, for that reason the first coefficient of all the polynomials will be 0 and this is not taken into account in the comparison, for mouth region three points have been used for the interpolation (the initial, the final and the center points), in eyebrow region four points are found to get the cubic polynomial that describes the region.

The comparison of the coefficients with the mean and the standard deviation of each coefficient in the polynomial is shown in Tables 2 and 3. The values of mean and standard deviation are near to zero, because of that control images and output images are similar.

In Fig. 4 results of our approach with some MMI database images are shown, the first column depicts to the input images. The second column are the output of EDEM algorithm and it can be seen that the noise in the images are due the illumination and in some cases the beard. In the ERED algorithm the noise is

Table 2. Mean of the error in the comparison of the polynomials of mouth and eyebrow regions.

	Coefficient 1	Coefficient 2	Coefficient 3
Lower mouth	0.101	.00070	X
Upper mouth	0.160	0.00104	X
Eyebrows	0.190	0.0016	0.000019

Table 3. Standard deviation of the error in the comparison of the polynomials if mouth and eyebrow regions.

	Coefficient 1	Coefficient 2	Coefficient 3
Lower mouth	0.105	.00071	X
Upper mouth	0.1518	0.001005	X
Eyebrows	0.177	0.0018	0.000035

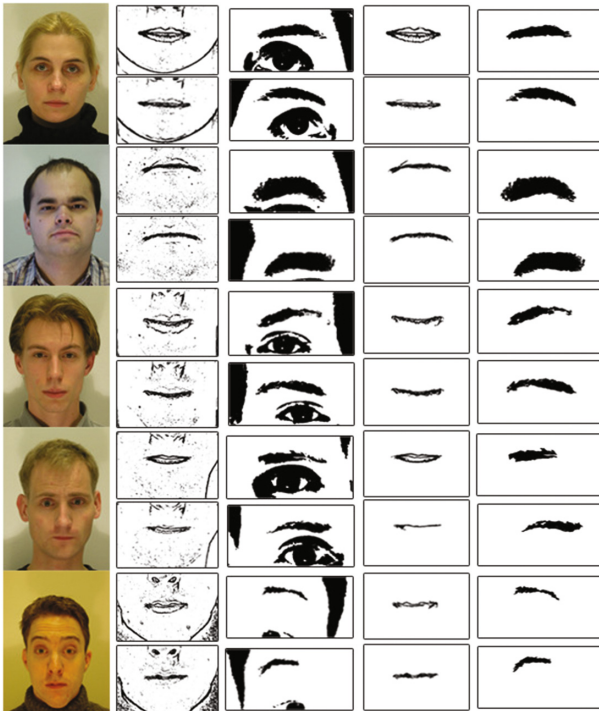


Fig. 4. Results of the MMI database: first column shows the input images; second column depicts the results from the EDEM algorithm in G and R channels; column 3 shows EDEM algorithm results; the last two columns show the DERBI algorithm segmentation.

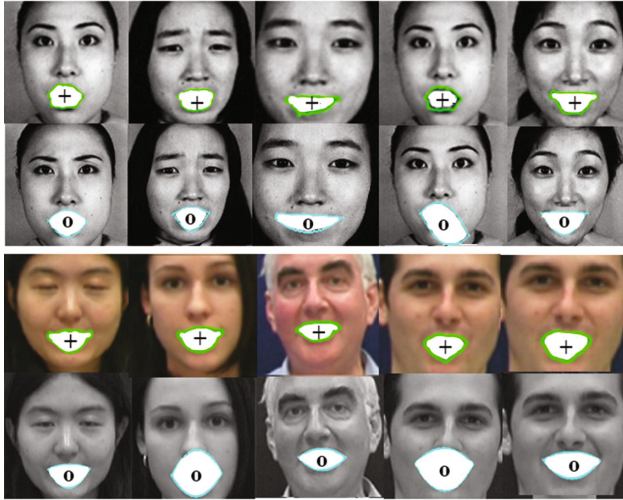


Fig. 5. Comparison of segmentation results, our results are marked by “o”, results obtained by [6] are marked by “+”.

related to eye region and the borders of the hair, the output images are depicted in the third column. Finally the images from the DERBI algorithm are shown in the last two columns. It can be seen that the found regions can describe the segmented ROI, and they could be used to find numeric features that define the region.

Additionally, we report a comparison with the results reported in [6] with jaffe and VidTIMIT databases. For this databases image equalization over mouth region it is applied, since these images have less quality than the MMI database and jaffe database was taken in gray scale. The mouth segmented region are indicated by the white area and the contour of the found region. Our results are depicted in Fig. 5 marked by “o”, results reported in [6] are marked with “+”. It can be seen that our approach is able to segment the mouth ROI in most of the cases. Particularly, the wrong segmentation results are due the illumination region in nose, this is because there is shadow between the nose and mouth, other reason is the result of the edge detector.

4 Conclusions and Future Work

In this paper a method to segment mouth and eyebrows is presented, the proposed method is based on transformations of the traditional regional and point filters to find the borders in mouth region and the eyebrows in face images.

Additionally, a method to find the ROIs in a binary image is proposed, this method is based on morphological operations and in DBScan algorithm to detect clusters in a binary image, then some metrics were used to detect the ROI in the image.

According to the experiments, our approach is able to segment mouth and eyebrows in most of the cases for the three datasets and the performance of our method is competitive when comparing with that proposed in [6].

As future work we are going to extract features from the segmented regions in order to either train or combine supervised learners for predicting expressions in face images.

Acknowledgments. This work was partially supported by the CONACyT Mastering Scholarship 701191, and the project OLLJ-ING17-I, VIEP-BUAP.

References

1. Corneau, C., Oliu, M., Cohn, J., Escalera, S.: Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applicants. *IEEE Trans. Pattern Anal. Mach. Intell.* **99**, 2–20 (2015)
2. Wenyi, Z., Rama, C., Phillips, P., Azriel, R.: Face recognition: a literature survey. *ACM Comput. Surv.* **35**(4), 399–458 (2003)
3. Lucey, S., Sridharan, S., Chandran, V.: Adaptive mouth segmentation using chromatic features. *Pattern Recogn. Lett.* **23**(11), 1293–1302 (2002)
4. Panning, A., Niese, R., Al-Hamadi, A., Michaelis, B.: A new adaptative approach for histogram based mouth segmentation. *Int. J. Electr. Comput. Energ. Electron. Commun. Eng.* **3**(8), 1564–1569 (2009)
5. Danielis, A., Giorgi, D., Larsson, M., Strömberg, T., Salvetti, O.: Lip segmentation based on Lambertian shadings and morphological operators for hyper-spectral images. *Pattern Recogn.* **63**(1), 355–370 (2017)
6. Le, H., Savvides, M.: A novel Shape Constrained Feature-based Active Contour model for lips/mouth segmentation in the wild. *Pattern Recogn.* **54**(1), 23–33 (2016)
7. Martins, P., César, F., Nardênio, A.: A real-time eyebrow segmentation and tracking technique to support an electric wheelchair interface. In: *Proceedings of International Conference on Computer as a Tool (EUROCON)*, pp. 1–6 (2015)
8. Hoang, L., Prabhu, U., Savvides, M.: A novel eyebrow segmentation and eyebrow shape-based identification. In: *Proceedings of IEEE International Joint Conference on Biometrics*, pp. 1–8 (2014)
9. Viola, P., Jones, P.: Rapid object detection using a boosted cascade of simple features. In: *Proceedings of Computer Vision and Pattern Recognition*, pp. 511–518 (2001)
10. Otsu, N.: A threshold selection method from gray-level histogram. *IEEE Trans. Syst. Man Cybernetics* **9**(1), 62–66 (1979)
11. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**(34), 226–231 (1996)
12. Pantic, M., Valstar, M., Rademaker, R., Maat, L.: Web-based database for facial expression Analysis. In: *Proceedings of International Conference on Multimedia and Expo*, pp. 5–10 (2005)
13. Michael, J., Shigeru, A., Miyuki, K., Jiro, G.: Coding facial expressions with gabor wavelets. In: *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 200–205 (1998)
14. Conrad, S., Brian, L.: Multi-region probabilistic histograms for robust and scalable identity inference. In: Tistarelli, M., Nixon, M.S. (eds.) *ICB 2009*. LNCS, vol. 5558, pp. 199–208. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-01793-3_21](https://doi.org/10.1007/978-3-642-01793-3_21)