Evaluation and Development of Strategies for Facial Features Extraction for Emotion Detection by Software

Sebastián Rincón-Montoya
and Carolina González-Restrepo
Mathematical Science Department
Mathematical Engineering
EAFIT University
Email: srincon2@eafit.edu.co
cgonza49@eafit.edu.co

Daniel Sierra-Sosa
and René Restrepo-Gómez
Optics Research Group
EAFIT University
Email: dsierras@eafit.edu.co
rrestre6@eafit.edu.co

Olga Lucía Quintero-Montoya
Mathematical Science Department
EAFIT University
Email: oquinte1@eafit.edu.co

***Abstract*—Nowadays the attention the emotion detection has catch around the world has lead to proper investigations.Trying to replicate the only validated emotion recognition system, FACS (Facial Action Coding System) the present investigation proposes a strategy for facial feature recognition process which is the previous step for an emotion detection, using as a base the canon of proportions, optical processing and mathematical engineering techniques. The results obtained with the proposed strategy are promising in terms of the diversity of faces that can be processed and the quality of the images, which suggests a further accurate emotion detection.**

Keywords: canon, detection, facial features, genetic algorithms, image processing, optical processing, Viola-Jones.

Figure 1: Universal emotions

1. Introduction

The understanding of the being and the control needs humans have over others has boosted the development of tools capable of deciphering people. Psychologists have been able to determine facial micro expressions lead to an emotion; the question that arises from the above is, can emotions be detected by software? This question has been addressed by others, using in the majority of cases, artificial intelligence techniques [1], also applying robotics to detect emotions [2].

Emotion detection is a process that aims to recognize and identify one of the six innate emotions, which are independent of culture, such as happiness, sadness, anger, fear, disgust and surprise [3], see Fig. 1. As technological advances increase more accuracy is expected in tasks performed by software; in emotion detection a precise recognition of an emotion can lead to a possible deception detection. As it is said in [4], a series of changes in emotions can show signs of a lie.

The present research is a continuing part of a previous research developed last year, oriented to the establishment of marks in specific facial features, which are considered crucial for a subsequent emotion detection. This facial features includes eyes, eyebrows, nose and mouth. The results obtained through the previous algorithm were precise for optimal images, considering this as images with good quality and lightening conditions, images containing optimal faces, which are faces were the facial features are easy to extract, these means people with beard, very light, dark skin tone, bangs or glasses were not processed correctly. Therefore the major objective in the present research is the strengthen of the previous algorithm, so a greater diversity of images can be processed. This implies the treatment of non optimal images (noisy images), with new strategies that involves optical processing and mathematical engineering techniques.

The algorithm proposed in this paper would be validated with FACS (Facial Action Coding System) [5]. The validation consists on a comparison between the positions of the facial feature marks located with FACS and the one located by the proposed algorithm. The data base FACS uses has low quality images in gray scale, with a diversity of people to evaluate and enlarge set of facial characteristics.

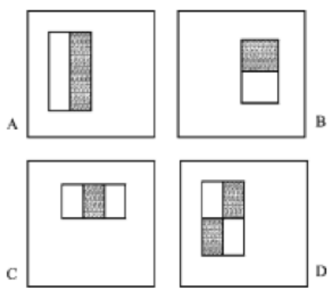
2. Theoretical Framework

2.1. Viola-Jones

Viola-Jones algorithm is one of the most used for real time detection, it is able to perform face and facial feature recognition in video or images. Its versatility allows detection of more than one face contained in the visual signal being used (images or videos). The accuracy of the algorithm lies in three pillars:

- Feature extraction
- Learning and classification algorithm
- Multi-scale detection

The algorithm detects faces classifying images based on simple feature values. The algorithm rather works with features than pixels, to simplify the learning process and to minimize computation time. The features used for detection involve the sum of pixel in the image within rectangles [6].



Viola-Jones features. Taken from [6]

The value of a feature v_f is given by,

$$v_f = \sum p_w - \sum p_d, \tag{1}$$

where p_w are the white pixels and p_d the dark pixels.

Once the feature value is known from (1), an "integral image" computes a value for each pixel (x, y) which represents the sum of the left and top pixels of the coordinate (x, y), including it. The speed of the algorithm emerge from this type of representation; it allows feature evaluation in constant time.

The classification consists on given a set of images containing recognizable objects, a classifier categorize that objects it distinguishes. The classifier is able to perform this task due to a previous learning process. Viola-Jones learning process not only trains the classifier but helps to choose the best feature.

Finally the algorithm has a multi-scale structure; it is a built cascade of classifiers that boosts the algorithm's ability of detection while reducing computation times. The cascade is made of different classifiers, each of them receives the set of images the previous classifier approved. This combination of techniques creates a facial detection system, which is substantially faster than any previous one.

2.2. Resizing Process

This process consists on readjusting the size of an image based on a certain criteria. In this specific case, Viola Jones detection algorithm is run through an image, giving as a result and image of a different size containing just the face the algorithm recognized. In order to standarize the measurements of the features needed to detect an emotion, a resizing process is used. All the faces recognized by Viola Jones are resized to 512×512 pixels. The chosen size is appropriate based on actual cameras' technology, which are in the worst scenarios 640×480 pixels.This process is also helpful to fixed parameters of certain used functions that depend strictly on the image size

2.3. Canon of Proportions

There are several facial features, such as eyes, nose, mouth and eyebrows that are vital for a micro-expressions recognition that leads to an emotion detection [7]. If the objective is an accurate emotion detection it is crucial to achieve a precise facial points' recognition.

There are conditions and proportions that fit in a natural concept and morphology of the human face, that allows a standardization call the canon of proportions. The canon may vary across cultures and through different ages, but the present work would be center in a Caucasian canon of proportions, see Fig. 2. The canon is a generalized model, this means that it is accurate but not an exact representation of the face.

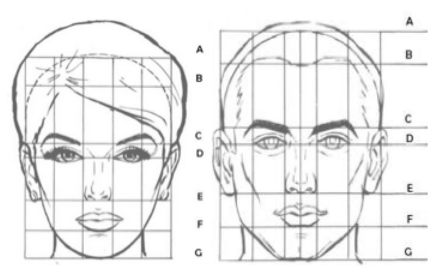


Figure 2: Canon of proportions

In the above figure the lines represent:

A: topline of the skull
B: hairline
C: eyebrow level
D: eyes level
E: bottom of the nose

F: bottom of the lip
G: bottom of the chin

The vertical lines sectioned the face in five equal parts from ear to ear. In terms of the measure of the canon, the width of the face must be three modules.

The studying of the canon of proportions helps the emotion detection problem, because it gives advantages in terms of mathematical proportions (module) of where should the facial features be. Some of the canon of proportions characteristics are listed ahead:

- The length of the face is three time the length of the forehead (module).
- The ears go from the eyebrows axis to the bottom of the nose. And they measure a module.
- The distance between eyebrows and pupils is a quarter of the module.
- The bottom of the lip is located in the middle of the bottom of the nose and the chin.
- The distance between eyes measures the same as one of them.
- The eyes are the center of the canon that goes from the topline of the skull to the chin.

2.4. Thresholding

Is a segmentation method. It is a nonlinear operation that converts a gray scale image into a binary image; this type of images has only two possible values for each pixel (0,1) [8]. This values or levels are assigned to pixels, which values are above or below a threshold value.

The purpose of using this operation in the algorithm is to reduce noise by filtering out unwanted pixels, with very low or very large values.

2.5. Adaptive Thresholding

Is a variation of the global thresholding, the difference is that in this procedure the threshold value is dynamic or local, this yields to better results, especially in images with varying levels of contrast. This procedure is performed with a convolution with a Gaussian window. The use of the adaptive threshold is vital to achieve accurate results in people with difficult or noisy features. This process reduce the noise without losing important information in the image. The parameter of this algorithm includes the size of the Gaussian window, which doesn't depend on the image size but the characteristics of the signal being processed [8]. Results of applying Thresholding and Adaptive Thresholding methods can be seen in Fig. 3.

Figure 3: Comparison using different thresholds. Taken from [8]

2.6. Shi-Tomasi Algorithm

It is a corner detector based on Harris corner detector but with some improvements. To understand Shi-Tomasi algorithm it is necessary to review the performance of the Harris detector:

There are certain features that are desirable in an image, and corners are one of them. A corner is just the intersection of two edges; this represents a variation in the gradient in an image. So to detect a corner you must identify a variation. Considering $w(x, y)$ a window moving through a gray scale image with displacements u in x-direction and v in y-direction. This window will calculate the variation of intensity I,

$$E(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2.$$

Whatever is in side parenthesis in w or I is the position is the window and the intensity in that position respectively. Consider the following expression,

$$\sum_{x,y} [I(x + u, y + v) - I(x, y)]^2.$$

If the above expression is maximized we can find windows with large variation of intensity. After using Taylor and expressing in a matrix form we obtained:

$$E(u, v) \approx \begin{bmatrix} u & v \end{bmatrix} M \begin{bmatrix} u \\ v \end{bmatrix},$$

where,

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}.$$

Then a value R is calculated for each pixel as following:

$$R = det(M) - k(trace(M))^2,$$

$$det(M) = \lambda_1 \lambda_2,$$

$$trace(M) = \lambda_1 + \lambda_2.$$

If R is greater than a certain value, then the window is a corner. The improvement Shi-Tomasi makes, is only the way the value R is calculated; for Shi-Tomasi R is the following:

$$R = min(\lambda_1, \lambda_2). \quad (2)$$

This slight difference creates big improvements applying (2). Even when Harris fails Shi Tomasi is able to detect corners.

Results of Shi-Tomasi algorithm can be seen in Fig. 4, where it shows the best corners in the image.

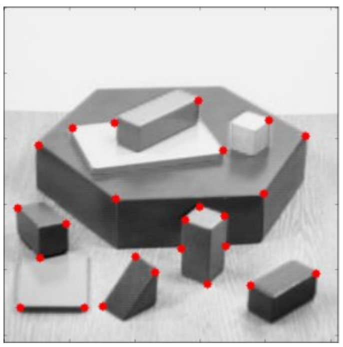


Figure 4: Results Shi-Tomasi. Taken from [9]

2.7. Hough Circle Transform

Circles are features of interest in an image, in our case to detect the pupils of a person. A circle of center (x_{center}, y_{center}) and radius r can be written as:

$$x = x_{center} + rcos(\theta), \quad (3)$$

$$y = y_{center} + rsen(\theta). \quad (4)$$

The points (x, y) from (3,4) describe the perimeter of the circle. Hough circle transform is used to determine parameters triplets as: $(x_{center}, y_{center}, r)$ to describe each circle which center falls in the perimeter above described. When the r is known the problem relapse in finding (x_{center}, y_{center}). This coordinates falls in a circle of radius r and center (x, y). The center (x, y) is common to all parameter triples, and can be found with Hough matrix or accumulation array.

2.8. FACS (Facial Action Coding System)

This coding system was first designed to taxonomize human movements [10]. It assigns each facial movement an Action Unit (AU), it can be interpreted as the smallest visible units of muscular activity in the face. For example, if a person moves his head to the left this Coding System assigns it a value of 51. Some possible combinations of these movements can be represented as the six universal facial emotions anger, disgust, fear, sadness, surprise, and happiness [11].

FACS has a large database, which will be necessary to validate the results obtained in this work. In order to validate the results obtained in this work, it is necessary to compare it with the database given by FACS, which is one of the most used coding systems for the facial expression analysis.

2.9. Genetic Algorithms

Genetic algorithms are an example of adaptive heuristic methods, which are commonly used to solve search and optimization problems. Genetic algorithms are based on genetic processes that living organisms follow [12]. Similarly to what happens in nature, genetic algorithms are based on a population of solutions, each of those solutions have a given value, representing individual chance to reproduce. The above implies the creation of better solutions, because the best individuals of the previous generation where chosen as fathers. The process leads to the convergence of the algorithm to optimal solutions, trying to avoid local solutions [13].

The use of heuristic methods in the present investigation is crucial to determine what combination of parameters must be used by the adaptive threshold in order to reduce noise and achieve a precisely facial features detection. The target of the genetic algorithm is to minimize the error between the real features and the marks established by the algorithm. The cost function that must be minimized is shown below:

$$E = \sum_{i=1}^{n} \sqrt[2]{(x - x_{est}))^2 + (y - y_{est})^2} \quad (5)$$

The genetic algorithm code is present below:

Algorithm 1 Pseudocode Genetic Algorithm

```
Read(Nparents,Nchildren,Ngenerations)
while Ngenerations and Convergence condition do
  parents=selection(generations)
  children=crossover(parents)
  if mutation=true then
    children=mutation(children)
  end if
  generation=parents + children
end while
solution=best(generation)
return solution
```

3. Methodology

To solve the feature detection problem for a further emotion detection, we first proposed an algorithm that involves image optical processing. The algorithm was developed with OpenCV (Open Source Computer Vision Library) in the Python interface, choosing Python as the programming language seemed correct since it is free to download and contains Viola Jones detection algorithm programmed for diverse features such as people with glasses. The flowchart of the proposed algorithm is shown in Fig. 5.

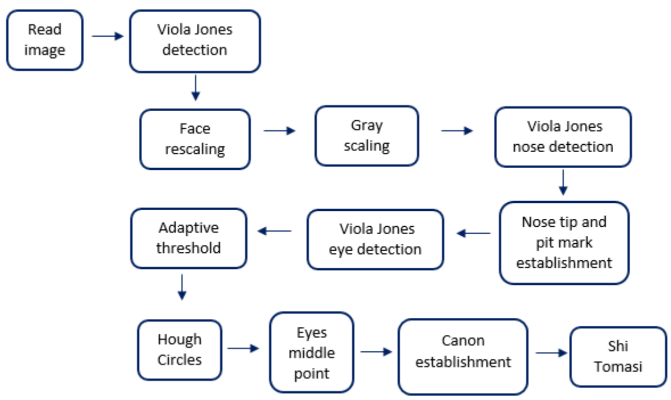


Figure 5: Flowchart of the proposed algorithm

The proposed algorithm uses four facial features (the two pupils, the tip and the pit of the nose) in order to establish the canon of proportions, the canon is establish even in non strictly frontal faces. The above is possible by calculating the middle point between the found pupils and the tip of the nose, the vertical line that joins this two points follows the inclination of the face contained in the image.

The results obtained with Python for high quality images were the desired. The unwanted characteristics in the face where removed and processed so that the marks established were in an accurate position, expecting this to improve the emotion detection process. The above proved that working with digital and optical processing was a correct option for feature detection and marks establishment, even in noisy images.

So what was the problem of the proposed algorithm? The major problem occurred when processing FACS images, which are necessary to validate the accuracy of the algorithm proposed in this paper. FACS images have low quality and poor lightening which affects the algorithm specifically the performance of the Viola Jones detection method, which is the basis of the proposed algorithm. If Viola Jones detection algorithm fails all the mark establishment process does to. To understand the magnitude of the problem it's necessary to review the major objective, which consisted on developing strategies in order to strengthen the past algorithm so a wider range of images could be processed, even low quality images such as FACS.

Seeking for a solution to the above problem we decided to translate the proposed algorithm to Matlab. The toolbox, Computer Vision System, from this package has a more robust Viola Jones algorithm, able to detect faces not only in high quality images. The transition to Matlab, strengthens the bases of the proposed algorithm, but there are other concerns that must be fixed. In order to the algorithm to be able to process great diversity of images, it must reduce and process the noise and the unwanted information contained in the image; this is achieved with the adaptive threshold. The results obtained with Python were precise largely due to the adaptive threshold parameters; which were adjusted manually for a each image; probing that the proposed algorithm lacks automation. For this reason the previous algorithm must be modified combining optical processing and mathematical engineering techniques, employing a genetic algorithm to determine the set of parameters the adaptive threshold must use. The set of parameters were found by testing a large set of FACS images with the genetic algorithm.

Matlab counts with the same functions used in the Python algorithm but they vary in terms of parameters, in order to achieve a more precise establishment in the eye brows and mouth, the parameters used by Shi-Tomasi should be optimize such as the ones in the adaptive threshold. For these reason the flowchart of the Matlab algorithm can be seen in Fig. 6.

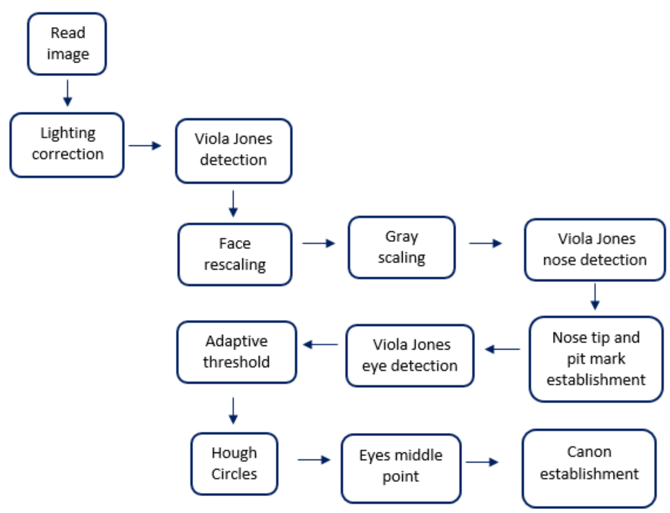


Figure 6: Flowchart of the proposed algorithm

The new algorithm includes an image lighting correction to improve the Viola Jones algorithm performance while processing the low quality FACS images.

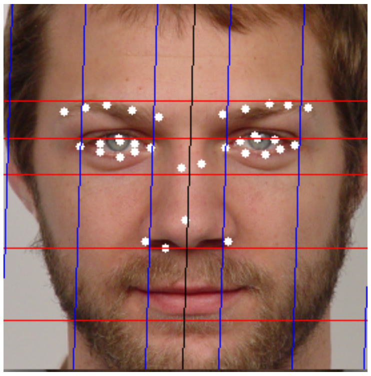


Figure 7: Noisy images in Python

Table 1: Genetic algorithm parameter estimation

Block size	Weighted mean	Threshold	Error
7	6.2272	0.3070	5.3423

4. Results

This section describes some of the results obtained with different methodologies that were used during the work, as it was explained in the previous section. As well as some of the difficulties that led to changes in the structure initially proposed.

4.1. Python results

It can be seen that the image in Fig. 7 can be considered as noisy because of the beard and the skin tone, which makes harder the establishment of the marks, despite all of this the algorithm could do it in a precise way.

It was because of the Adaptive Threshold that what we know as noise and some unnecessary information can be eliminated.

4.2. Genetic Algorithms

With the need to generalize the parameters that will work best for the proposed algorithm, it was designed this heuristic for the estimation and we used 100 images from FACS database.

The parameters use in the Genetic Algorithm to obtain the results in Table 1 were: $N_{parents} = 30$, $N_{child} = 50$ and finally $N_{generations} = 15$. The error from Table 1 was obtained from (5), calculating the errors between the real values (given by FACS) and the estimated.

4.3. Proposed algorithm results

It can be seen in Figs. 8 and 9 how the algorithm is also able to establish the necessary marks for the construction of the canon even for images with low resolution, poor lighting and noise.

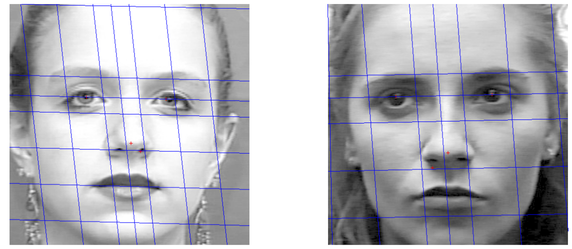


Figure 8: Canon establishment

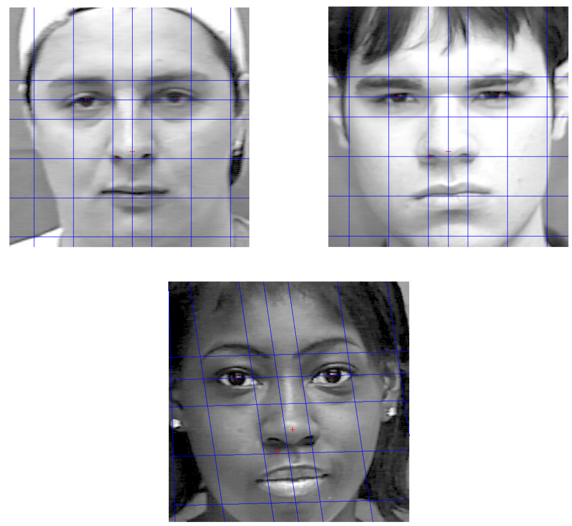


Figure 9: Canon establishment for noisy images

It remains to implement the genetic algorithm to estimate the parameters of the function used in Python to detect the remaining feature points (Shi-Tomasi).

5. Conclusions

The result obtained in Python fulfills the objective of the present investigation; where more marks were established, even in a noisy image, giving the more information about the facial features which facilitates the detection of an emotion.

It is clearly the vitality of the Viola Jones detection algorithm to segment the face facilitating the feature mark establishment; in the same way the establishment of the canon of proportions is crucial for a further emotion detection, by analyzing the changes in position of the facial features from a neutral face to a face with an emotion.

The value of the parameters obtained for the adaptive threshold, with the genetic algorithm seemed to be the expected ones. These is evidenced in the results obtained were the marks are established correctly even in noisy images. This probes the power of the adaptive threshold to filter noise and unwanted information.

The way of establishment of the canon allows the processing of non frontal faces,which enlarges the set of images that can be processed by the algorithm.

The lighting correction improves the performance of Viola-Jones, even though several images couldn't be processed by Viola Jones detection algorithm. The evaluation different detection algorithms is recommended in order to strengthen the performance of the proposed algorithm. A more robust marks establishment algorithm is expected when the parameters of the other functions employed can be optimized.

References

[1] Y. Zhang, L. Zhang, and M. A. Hossain, "Adaptive 3d facial action intensity estimation and emotion recognition," *Expert Systems with Applications*, 2014.

[2] L. Zhanga, M. Jiangb, D. Faridc, and M. Hossaina, "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot," *Expert Systems with Applications*, vol. 40, p. 5160–5168, Oct. 2013.

[3] NeuroWikia, "Bases neurobiólogicas de las emociones," 2013.

[4] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[5] P. Ekman and W. V. Friesen, "Facial action coding system," 1977.

[6] M. Jones and P. Viola, "Fast multi-view face detection," *Mitsubishi Electric Research Lab TR-20003-96*, vol. 3, p. 14, 2003.

[7] B. Hogarth, "Drawing the human head," 2011.

[8] O. python tutroals.readthedocs.org, "Image thresholding — opencv-python tutorials 1 documentation," 2015.

[9] O. python tutroals.readthedocs.org, "Harris corner detection — opencv-python tutorials 1 documentation," 2015.

[10] C.-H. Hjortsjö, *Man's face and mimic language*. Studen litteratur, 1969.

[11] P. Ekman and E. L. Rosenberg, *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, 1997.

[12] C. Darwin, "The origin of species by means of natural selection, or, the preservation of favored races in the struggle for life. vol. 2," *International Science Library*, 1859.

[13] M. Mitchell, *An introduction to genetic algorithms*. MIT press, 1998.