

# Facial Recognition with Mobile Application and Artificial Neural Network

Glenda Botelho<sup>1</sup>, Marcelo Cobias<sup>1</sup>, Anna Paula Rodrigues<sup>1</sup>, Ivo Sócrates M. Oliveira<sup>2</sup>, Maxwell Diógenes Bandeira Melo<sup>1</sup>, Ary Henrique M. Oliveira<sup>1</sup>

<sup>1</sup>Computer Science Department, Federal University of Tocantins, Palmas, Tocantins, Brazil

Email: <sup>2</sup>Computer Science Department, Federal Institute of Tocantins, Palmas, Tocantins, Brazil

Email:ivo@ifto.edu.br

**Abstract**—This work presents the implementation of a facial recognition system with mobile application for the identification of people/faces. For this, haar-like object detection techniques, filters of luminosity, contrast and grayscale are used. In order to abstract the characteristics, the discrete cosine transform (DCT) and the Laplacian filter were used. In the classification stage it was used the Multi-Layer Perceptron Neural Network (MLP) and Self-Organizing Map (SOM). The interaction flow between the application steps and the classifier has been linked to a web services set. The results reached an accuracy of up to 97%, reaching the objectives proposed for the work.

**Keywords**—facial identification; neural networks; image processing; mobile app; features extractor; facial segments.

## I. INTRODUCTION

Falsification of identity is a crime against the public faith with the intention to gain an advantage over a third party to result in a gain or cause harm [1]. To avoid this crime, different computational techniques are proposed. One particular technique is biometrics, which aims to extract and define characteristics of an individual in a way that makes it unique, or in other words, more easily identifiable. In this context, one aspect of biometrics is facial recognition, which uses knowledge in the area of artificial intelligence, computer vision and image processing.

Facial recognition can be defined as a technique to identify patterns in physical characteristics such as mouth shape, face, distance of the eyes, etc[2]. The human being recognizes easily a family person, even with obstacles preventing their perfect vision. However, for a machine, this process is not trivial, requiring multiple procedures to detect and recognize specific patterns capable of labeling a face as a known or unknown, for example.

Unlike other biometrics models, facial recognition does not require the use of specialized equipment and can use simple hardware (mobile cameras, for example), allowing

the identification of more than one individual simultaneously in a single unit. Thus, in order to take advantage of such features, this paper presents the development of a personal recognition system using a mobile application for support.

Thus, a few steps were taken, such as personal image capturing by the device, characteristic segments extraction (face, mouth, eyes and nose), the segments normalization through filters, significant features extraction in segments found and the segments classification in relation to the training images found on the base.

## II. RELATED WORKS

Several facial identification models can be found in the literature. Each variant differs according to the approach type to detection, extraction, classification features, besides the application.

In [3], is proposes a face recognition method based on higher order statistics (HOS) applied to public security. This work aimed to identify individuals with criminal bond, previously registered in a database. HOS is used to create compact face signatures in addition to Fisher's Discriminant Ratio (FDR) and linear correlation to eliminate redundancies. The results showed a detection and classification rate above 70%.

In the work of [4], a multi-purpose algorithm was implemented for: face detection, face alignment, pose estimation, gender recognition, smile detection, age estimation and facial recognition, simultaneously, using a simple deep convolutional neural network. Because it is a multitasking problem, it was necessary to use a learning framework in order to facilitate synergy between different domains and application tasks. According to the authors, several experiments have shown that such networks presented better results for understanding faces and achieved satisfactory results for most tasks.

In [5] is shows a three factors authentication system of face recognition, gestures and location. Due to the users gestures and location being time series, the authors used a

recurrent LSTM (Long Short-Term Memory) type with unsupervised learning. The work has promising partial results, demonstrating the method viability.

Finally, [6] uses a Support Vector Machine (SVM) to classify human emotions. Facial expressions recognition to extract human emotions is a growing field in computer vision. In this work, the proposed system combines the cloud model with the traditional model. Facial Landmarks and Center of Gravity (COG) algorithms are used, which generate training and test data sets that contain expressions of anger, disgust, fear, happiness, neutrality, sadness and surprise. The proposed system was tested on CK+, JAFFE and KDEF databases, reaching a prediction rate of 96.3%.

### III. METHODOLOGY

For this work, it was proposed a system implementation composed of a mobile application integrated with a facial identification method. This structure can be divided into three parts: a mobile application, responsible for capturing the faces and presenting the results of the captured image identification; a communication channel between web services that organize the personals basis to be registered and establish communication between the application and the identification method; and an identification method, responsible for centralizing the personal base registered and transform into a learning base.

The segmentation process corresponds to the stages of normalization and detection of facial segments characteristic. The segmentation criteria used in this paper correspond to the morphology facial nature. The proposed segmentation methodology consists of samples normalization through luminance, contrast and color filters and the facial segments detection such as face, eyes, mouth and nose through classifiers based on Haar-like features [7]. The segmentation phase is very important in the identification process and should be as efficient as possible, for lack of accuracy undertake the subsequent processes of identification.

For comparative purposes, two identification system methodologies were modeled. Both methods process the extracted face region by the Haar cascade detector. The first method extracts the most significant pixels of the face by a DCT (Discrete Cosine Transform), in order to reduce its dimensions to feed a multilayer perceptron network, which classifies the individual. The second model extracts the segments characteristic shapes using a Laplacian filter, and identifies the individual using networks of self-organizing maps.

#### 2.1. Mobile Application

The mobile application acts as a terminal interaction between the user and identification systems. This

application allows the user to do three operations: capturing image from the device camera, checks that there is a face in the image, and query the identity of the individual by the verification system. No significant transaction or processing charge is made in this application, aimed at saving of mobile resources and to allow a more fluid running even on devices with more modest hardware.

#### 2.2. Web Services

In order to communicate with the mobile application, a set of web services has been implemented: a public, which provides operations and data that do not change the identification data set; and a private one, which implements the personal registration operations, as well as others that influence the system recognition capacity.

Images of person registered by the web service will be stored and indexed in the database. These are handled by the identification system with the normalization and segmentation processes for the facial features extraction. Later they are transformed into knowledge base for identification.

#### 2.3. Identification System

In this paper, a set of steps were implemented to process and generate knowledge for identification. These steps are presented in the Fig 1.

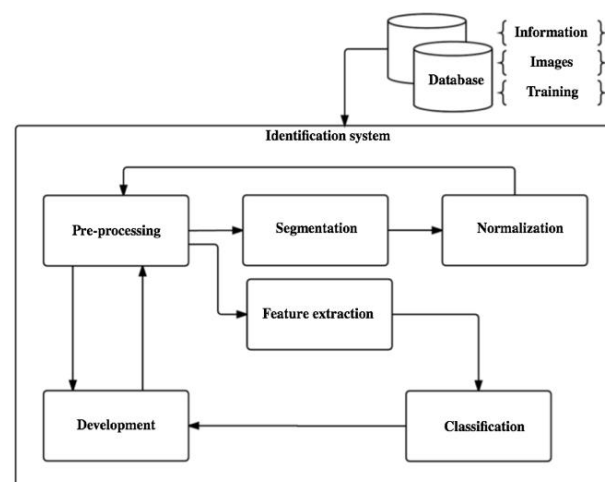


Fig. 1: Identification system and its respective operations.

The first step of the network input processing consists in image normalizing. This step aims to improve the morphological aspects in order to increase the chances of success of the following steps [8]. To realize the normalization was applied three filters: grayscale, luminance and contrast. Grayscale filter converts the images in RGB format to grey scale. The luminance filter adjusts the brightness intensity using linear filters. Finally, the contrast filter uses the histogram equalization technique resulting in constant levels of brightness in each pixel of the image. The result of the application of

the filters can be seen in the Fig2, where they are presented by the original image and adjustment in grayscale, luminance and contrast, respectively.



Fig. 2: Normalization filters. (a) Grayscale (b) Luminance (c) Contrast.

After normalization, the segmentation stage consist in the extraction of facial characteristics like eyes, mouth and nose [9] by classifiers based in Haar-like features [7]. In the proposed segmentation stage was used four public haar-like representation models for segmentation. Such models are presented in the Table 1.

Table. 1: Public haar-like models used.

| Model        | Size  | Author              |
|--------------|-------|---------------------|
| Mouth        | 25x15 | Santana et al. [10] |
| Nose         | 25x15 | Santana et al. [10] |
| Front Face   | 24x24 | Lienhart et al. [9] |
| Pair of eyes | 45x11 | Santana et al. [10] |

The Fig3 shows the operations performed in the segmentation process: quantization, detection and cleavage.

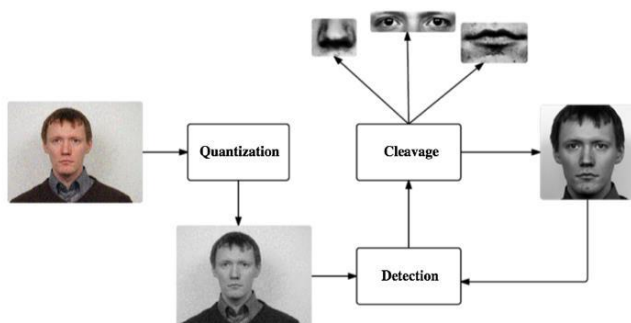


Fig. 3: Normalization filters. (a) Grayscale (b) Luminance (c) Contrast.

- Quantization: performs the image preparation process, that is, adjustment of the pixels intensity variation present in an image. The lower the pixels intensity present in an image, faster and accurate will be the detection operation. For the quantization, 8 bits was used per pixel intensity;
- Detection: in the first instance the detection uses a type of haar-like model for looking for a specific image area. This model corresponds to the delimitation of an object similar to a human face. After we found the segment by detecting, the same will suffer the process of cleavage;

- Cleavage: physically delimits a specific part, that is, performs the division of certain characteristics. To find the other characteristic segments, the face cleaved of the image suffer another detection process, using models haar-like corresponding to the pair of eyes, mouth and nose. At the end of the process, characteristic facial segments will be highlighted from the input image.

For the characteristic extraction step, two different methodologies were modeled, which are the image compression by transform DCT and the application of laplacians filters for edge detection. Two-dimensional DCTs were used to extract the image DC coefficients, along with the application of a ratio between the extracted coefficients and a highlight matrix (luminance or chrominance). Each DCT coefficient was mapped to a finite levels number determined by the compression factor. Compression factors are defined by the subdivided blocks number for DCT application ((4x4); (8x8); (16x16); (32x32); etc) and the quality factor defined by an enhancement matrix. At the end of the method the IDCT(Inverse Discrete Cosine Transform) was applied for image reconstruction. In this work was applied 8x8 blocks and the matrix used for the quality factor was the luminance, in which defines the color spectrum levels resulting in the image.

The second method used a Laplacian and morphological filters. The combination of these filters aims to highlight contours and edges, which correspond to the facial features extracted from the segmentation stage. The morphological filter highlights the edges obtained by the Laplacian filter. Through the image-opening operation, the noise will be removed and the edges found by the Laplacian filter will be highlighted.

Finally, in the classification stage, two models were used: one supervised and one unsupervised. The supervised model is represented by a MLP network with Resilient Back-Propagation (Rprop) learning algorithm. This methodology was based on the method applied by [11] on facial recognition based on neural networks combined with transform in the images domain. Thus, the input samples for learning the MLP network go through a characteristic extraction step based on the discrete cosine transform (DCT). This step serves to represent the more compact image, to reduce the amount of computational effort required for training and classification of MLP.

Already proposed unsupervised network is represented by a SOM based on the competitive learning method. This methodology was based on a concept published by [12] on the use of the SOM classifier to optically recognize certain types of characters. Thus, the input samples for the SOM network use the sum of Laplacian and

morphological filters to extract forms belonging to the facial segments.

At the end of the classification the set of characteristic segments (face, mouth, eyes, nose) was applied to training. The training process resulted in four respective knowledge bases to the sets extracted from the previous stages of the identification system. The knowledge generated was stored in a database, it will be later read to the identification phase carried out by the application.

#### IV. RESULTS AND DISCUSSION

In this work a facial identification system was implemented based on two classification methodologies: MLP with DCT and SOM with Laplacian filter. Both using the basis of facial images of Denmark and

Nijmegen universities. The Denmark is composed only of male individuals. Despite being a relatively small base, the samples have a nice quality. Already, the Nijmegen base has a lower quality than Denmark, but has a large variety of individuals in relation to gender, age and ethnicity.

##### 4.1. MLP classification methodology with DCT

For this methodology a total of 42 tests were performed, varying the number of neurons in the hidden layers, the number of hidden layers, the size of the images, the error and the number of cycles. For one hidden layer the results are presented in the Fig. 4 and for two hidden layers in the Fig. 5, both for the Denmark database. For the Nijmegen database, for one hidden layer the results are shown in the Fig. 6 and for two hidden layers in the Fig. 7

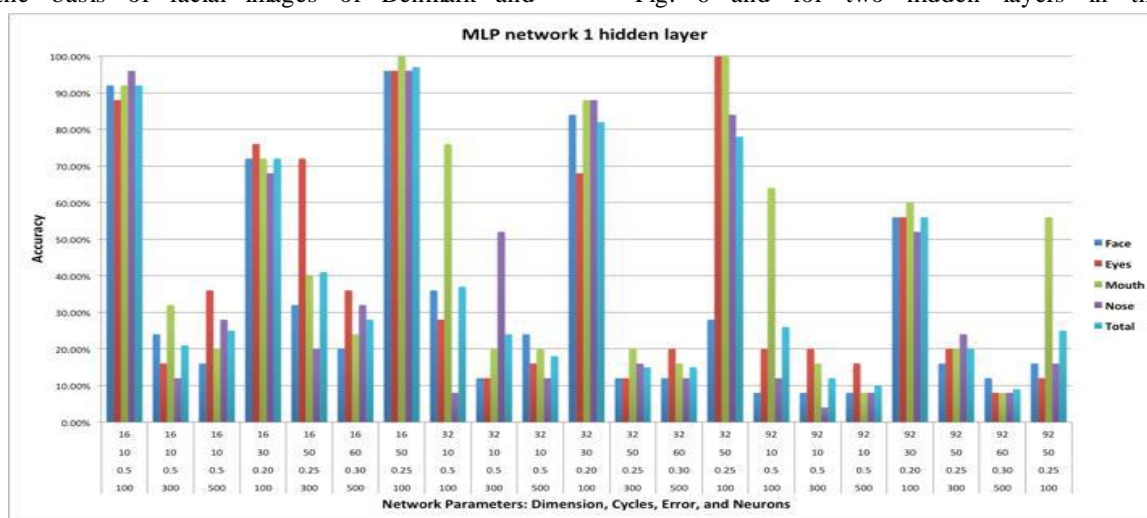


Fig. 4: Result of one hidden layer MLP network with the Denmark database.

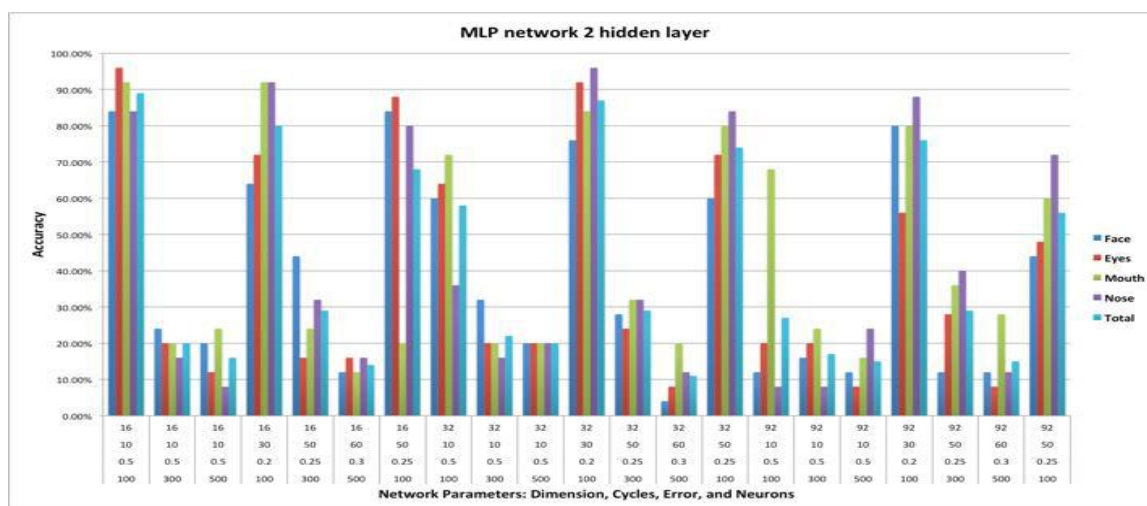


Fig. 5: Result of two hidden layer MLP network with the Denmark database.



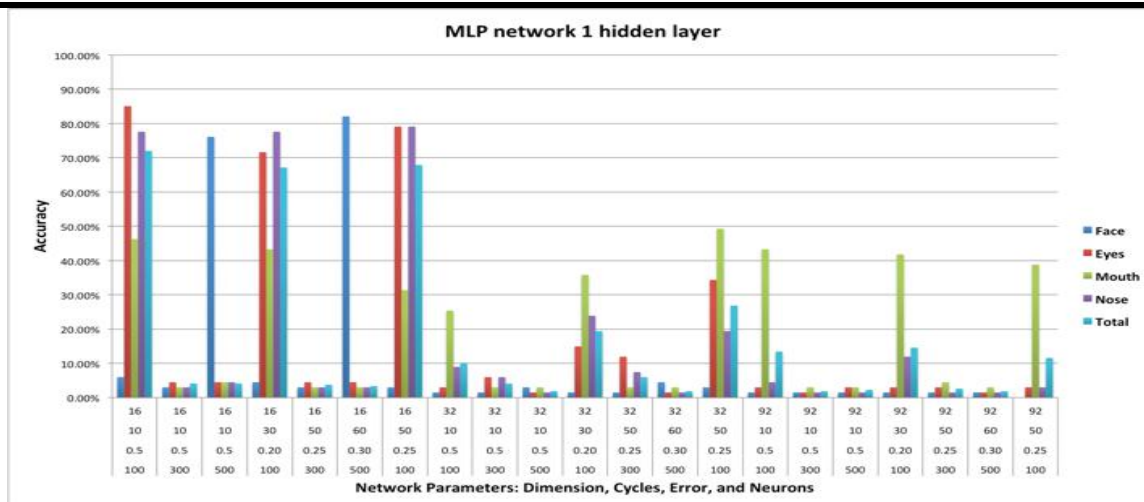


Fig. 6: Result of one hidden layer MLP network with the Nijmegen database.

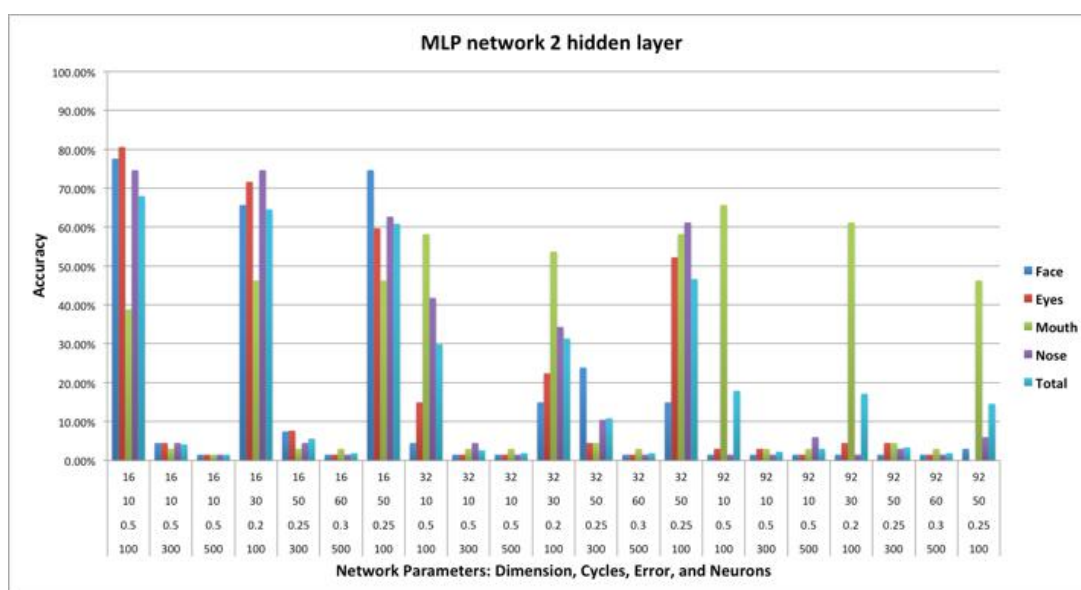


Fig. 7: Result of two hidden layer MLP network with the Nijmegen database.

It can be seen that the best results were observed in Denmark database, with the network composed of one hidden layer, with smaller amounts of neurons per layer and a minimum error rate ranging from 0.20 to 0.30. The best accuracy was obtained with the configuration: 100 neurons per layer, one hidden layer, an error of 0.25 and 50 cycles. For this configuration, the accuracy rates by characteristic segment were: 96% for the face, eye and nose, 100% for mouth and 97% of the total.

For the images obtained from the Nijmegen database, it is noted that the best results corresponded to the parameter of 100 neurons per layer and a minimum error rate that ranged from 0.5 to 0.25. As for the Denmark base, the best result had one hidden layer. The hit rate for this best result was 72,02%, where the eye characteristic segment contained 85,07% hit and the others had a margin of error of 33%.

Among the set of performed experiments some parameters achieved success. The image size 16x16 showed better results compared to other dimensions. One hundred (100) neurons per layer obtained the best accuracy in identifying the characteristic segments. The minimum error and cycle parameters obtained good adjustment intervals, ranging from 0.5 to 0.25 in the case of the minimum error and 10 to 50 in the case of the cycle.

#### 4.2. SOM classification methodology with Laplaciano filter

For the SOM network methodology with Laplacian filter, three tests were performed varying according to the input images dimensions. Because it is an unsupervised network, it defines the number of neurons through the input vector (image dimension) and balances their weights randomly. In test 1 was used an input vector with 16x16 pixels, the test 2 an input vector with 32x32 pixels

and in test 3, a vector with 92x92 pixels. For the Denmark database, the results are presented in the Table 2 and for the Nijmegen database in the Table 3.

*Table. 2: Results of tests 1, 2 and 3 of the SOM network for the Denmark database.*

| Test Number | Face       | Eyes       | Mouth      | Nose       | Total      |
|-------------|------------|------------|------------|------------|------------|
| Test 1      | 40.00<br>% | 52.00<br>% | 28.00<br>% | 36.00<br>% | 39.00<br>% |
| Test 2      | 60.00<br>% | 56.00<br>% | 48.00<br>% | 40.00<br>% | 51.00<br>% |
| Test 3      | 56.00<br>% | 56.00<br>% | 48.00<br>% | 60.00<br>% | 55.00<br>% |

Analyzing the data in Table 1 we can see that the thresholds belonging hit rates did not achieve good results for identification. This may indicate that the network training was not satisfactory, leading to stagnation. Regardless of the input vector (size of images) the characteristic segments obtained low success. The test 3 showed the best hit rate, 55%, where the nose segment contained 60% hit and the other segments had a margin error of 56%.

In the Table 3 for the Nijmegen database, it is also observed that the thresholds belonging to the hit rates did not reach expressive results for the identification, which leads us to deduce that the training was also not enough. The test 2 can be considered the most successful, where the hit rate has reached 51.92%, the mouth segment with 58.21% hit and the other segments with a margin error of 49%.

*Table. 3: Results of tests 1, 2 and 3 of the SOM network for the Nijmegen database.*

| Test Number | Face       | Eyes       | Mouth      | Nose       | Total      |
|-------------|------------|------------|------------|------------|------------|
| Test 1      | 31.34<br>% | 25.37<br>% | 38.81<br>% | 34.33<br>% | 32.46<br>% |
| Test 2      | 46.27<br>% | 50.75<br>% | 58.21<br>% | 52.24<br>% | 51.86<br>% |
| Test 3      | 49.25<br>% | 53.73<br>% | 56.72<br>% | 46.27<br>% | 51.49<br>% |

#### 4.3. Quality Metrics

In order to analyze the effectiveness of the classification models used in this work, metrics were used to calculate the recognition quality rates. The first metric used was the False Recognition Rate (FRR). The FRR described in the Equation 1, is an error measure that indicates the percentage of individuals present in the knowledge base, which are not recognized by the classifier. The lower the FRR rate, the higher the system hit rate.

$$\text{FRR} = (\text{Number of False Rejections} / \text{Total Number of Recognized Accesses}) \times 100\% \quad (1)$$

In addition to this, the False Acceptance Rate (FAR), Equation 2, was also calculated. The False Acceptance Rate (FAR) presents the individuals percentage not known by the learning base, which are presented as existing by the classifier. The lower the FAR the higher the system hit rate.

$$\text{FAR} = (\text{Number of False Acceptances} / \text{Total Number of Unknown Accesses}) \times 100\% \quad (2)$$

Finally, the Total Success Rate (TSR) was calculated, Equation 3, which presents an overall performance measure combining the FRR and FAR error measures to extract the classifier identification rate.

$$\text{TSR} = [1 - (\text{FAR} + \text{FRR} / \text{Total Accesses Number})] \times 100\% \quad (3)$$

The three proposed metrics were used in the best results obtained in the test phase. For the FAR calculation, individuals from the Denmark database were used to be identified in the training created by the Nijmegen database, and 25 individuals from the Nijmegen database to be identified by training created from the Denmark database.

The Table 4 shows the proposed metric rates for approach cited. It is possible to note that the MLP classifier appeared to have some superiority in relation to the SOM methodology due to the low rates of FRR and FAR. This performance difference can be explained due to its nature of learning, while the MLP is supervised by means of more than one adjustment parameter, the SOM, because it is unsupervised, the network itself tries to understand the input parameters and organize them for recognition.

*Table. 4: Quality metrics applied to proposed methodologies.*

|     | FRR   | FAR   | TSR   |
|-----|-------|-------|-------|
| MLP | 26.86 | 9.25  | 73.05 |
| MLP | 4.00  | 2.00  | 88.00 |
| SOM | 47.76 | 32.90 | 39.80 |
| SOM | 44.00 | 29.00 | 46.00 |

#### V. CONCLUSION

This paper presented the development of a facial recognition system using mobile application. Thus, two classification methodologies were implemented, one of the MLP network with DCT and the other the SOM network with Laplacian filter. In addition, Denmark and Nijmegen facial image database were adopted.

Based on the results it is possible to conclude that in the classification stage the MLP network obtained better results when compared to the SOM network, reaching an accuracy of 97%. In addition, the SOM network presents FRR and FRA rates very closely demonstrating that the classifier will try to recognize an unknown

element. The characteristic segments corresponding to the nose and mouth, in both classification methodologies, obtained the best recognition rates.

It was concluded that the MLP network obtained the best results with the lowest input vectors (16x16) and with only one hidden layer. This can be explained as the more neurons in the hidden layer, the higher the convergence time and the probability of network stagnant. As for the SOM network, the topographic map of the same did not balance the distribution of the classes belonging to the input samples between the neurons.

The results were satisfactory, especially if we consider the difficulty of finding bases with facial images of high resolution. Most of these bases are found compacted, which makes the limited study. A greater number of samples could expand the results by obtaining a greater variety of interesting patterns, achieving a greater class of features.

#### ACKNOWLEDGEMENTS

To Brazilians research funding agencies FINEP and Research Program for Brazilian Single Health System (PPSUS) for financing the equipment used in the development of this work.

#### REFERENCES

- [1] Nucci G. S. (2007). Código do direito penal comentado (7th ed.), Editora Revista do Tribunais.
- [2] Handbook of Biometrics (2018). Retrieved from <https://www.researchgate.net/publication/321619782-Handbook-of-Biometrics>
- [3] da Silva Neto J. G., Caldeira J. L. M., Ferreira D. D. (2018). Face Recognition based on Higher-Order Statistics. IEEE Latin America Transactions, 16 (5), pp. 1508-1515.
- [4] Ranjan R., Sankaranarayanan S., Castillo C. D., Chellappa R. (2017). An All-In-One Convolutional Neural Network for Face Analysis. IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 17-24.
- [5] Rexha B., Shala G., Valon X. (2018). Increasing Trustworthiness of Face Authentication in Mobile Devices by Modeling Gesture Behavior and Location Using Neural Networks. Future Internet, 10 (17).
- [6] Alshamsi H., Kepuska V., Meng H. (2017). Automated facial expression recognition app development on smart phones using cloud computing. IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON) pp. 577-583.
- [7] Viola P. & Michael J. (2001). Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 1st ed.
- [8] Batista S. M. L. et al. (2013). Processamento digital de imagens para detecção e classificação de nódulos em mamografias. Revista Eletrônica da Faculdade Metodista Granbery, pp. 4-9.
- [9] Lienhart R., Kuranov E., Pisarevsky V. (2003). Empirical analysis of detection cascades of boosted classifiers for rapid object detection. Joint Pattern Recognition Symposium, pp. 297-304.
- [10] Santana C. S., Deniz O., Guerra C., Hernandez M. (2007). ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. Journal of visual communication and image representation pp. 130-140.
- [11] Risk M. R. N. and Taha A. (2002). Analysis of neural networks for face recognition systems with feature extraction to develop an eye localization based method. Electronics, Circuits and Systems, pp. 847-850.
- [12] Marinai S., Marino E. and Soda G. (2007). Word retrieval in document images without OCR, Proceedings of the Eleventh Italian Symposium on Advanced Database Systems.