Designing and implementation of MACE and HBCOM filters for facial verification via a computational simulation of Vander Lugt's optical correlator.

Diego A. Herrera Rojas*[1], Luis C. Duran Neme†[1], Andrés F. Duque Bran‡[1], and Nicolás A. Ávila Pérez§[1]

[1]Physics Department, Science Faculty, National University of Colombia - Bogotá, Optics and Acoustics

July 12th of 2020

Abstract

The main purpose of this project is to explore correlation techniques for facial recognition focusing on the implementation of MACE and HBCOM linear filters as well as on the Vander Lugt's optical correlator from a digital image-processing perspective. With this aim, contents related with database acquisition, preprocessing techniques, filter synthesizing, normalization metrics and ROC curves analysis are necessarily revised through this article. The algorithms were implemented on four distinct subject databases with an accuracy greater than 80 % in each case.

Key words. Facial recognition, optical correlators, linear filters, Vander Lugt, normalization metrics.

1 Introduction

This article studies correlation techniques for facial recognition, exploring the capabilities of such techniques as a non-segmentation alternative for biometric recognition. Particularly, a digital implementation of correlation is considered as an illustration of the scope of these methods. A simple application that performs identity verification and a simple threshold-based decision algorithm are developed using MATLAB. Therefore, the main goal is to design and implement a correlation-based, static facial verification algorithm. That is, an algorithm capable of automating the recognition of a person using their face, with an accuracy higher than 80%, which is done by performing a facial verification process based on a purely optical fashion represented by the computational simulation of a Vander Lugt correlator as well as the inclusion of the MACE and HBCOM linear filters.

To achieve this objective, is necessary to determine an image database acquisition protocol that allows synthesis of a robust linear filter in terms of intensity and noise in the image, as well as the development of an acquisition routine that automates this process. Additionally, the designing and synthesizing of the filters shall be done by implementing a robust algorithm both for MACE and HBCOM filters. Then, by means of normalization metrics, that allow correlation peak characterization, and the performance a ROC curve analysis, parameters of the decision are chosen to achieve the proposed confidence range (higher than 80%) in the performance of the application under ideal conditions of illumination.

1.1 Optical Correlators

Historically, correlation filters have been implemented using optical correlators based on the 4f configuration, a JTC configuration or a Van-

*dieaherreraroj@unal.edu.co

†lcdurann@unal.edu.co

‡afduquebr@unal.edu.co

§navilap@unal.edu.co

der Lugt 4f configuration. The essential principle is the capacity of lenses to compute the Fourier Transform of a diffraction pattern in its back focal plane. Usually, the optical set up is complemented with optoelectronical devices capable of storing several correlation filters for biometric recognition, which when located at the Fourier plane, are correlated with a query image using a second lens. The usage of correlation techniques has been largely developed in the context of optical correlators and , as mentioned before, the 4f set up is the most commonly implemented. Generally, the physical montage for these correlators is based on the scheme shown in figure (1)

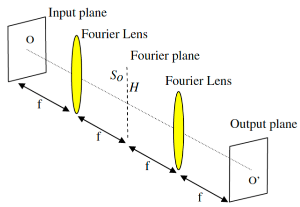


Figure 1: Basics for the optical correlators physical montage [8].

As an additional note, is worth mentioning that the JTC architecture has been used in some applications on the facial recognition field, with different setting variations. However, the main focus of the project will be on the VLC correlator which is described bellow.

1.1.1 Vander Lugt Correlator

The Vander Lugt Correlator or VLC, just as the JTC, is based on a "4f" setup and uses the spectrum of target and reference images in order to dictates how similar they are by optically computing and displaying the correlation. Nevertheless, the VLC settings and performance features differ drastically from the JTC, mainly because in the former the input plane only contains the target image.
Perhaps the most important characteristic of the VLC architecture is the usage of correlation filters, which are usually constructed from a group of reference or training images and are meant to be applied to the test image spectrum. Afterwards, a second Fourier transform is computed and finally the result appears in the correlation plane as a single peak, which ideally, must be sharp, bright and as centered as possible. The basics of how the VLC works are presented in the diagram on figure (2).

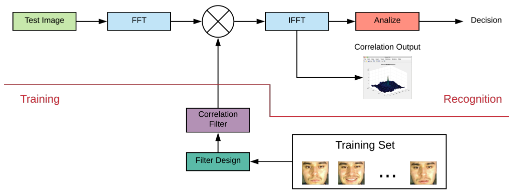


Figure 2: Diagram for VCL correlator working process.

For the physical montage, the idea is to use a linearly polarized parallel beam as a light source to illumine the input plane, after which the test image FT is performed thanks to a lens, next the chosen filter is applied employing a SLM modulator on the Fourier plane. Finally, a second lens performs yet another FT and the resulting correlation plane is to be displayed on a CCD camera [1].
As was mentioned before, the correlation filters play a huge role when it comes to the VLC architecture, therefore is necessary to construct an appropriate filter to obtain the most favourable result possible.

1.2 Linear Filters

1.2.1 MACE

Linear filters are computed by forming linear combinations of training images in frequency space. Although this sounds simple, the characteristics of the correlation output can be controlled by choosing the superposition coefficients so that a particular property is optimized. That is the particular case of linear filters like the MACE. The MACE Filter **minimizes the Average Correlation Energy** of the correlation output (hence the name). But also, this particular filter is designed so that the correlation output has a fixed peak size for the training images used. So, the MACE filter is a constrained linear filter.
When training images are viewed as column vectors $\mathbf{x_i}$, and the MACE correlation filter, as a column vector $\mathbf{h}$, the correlation output of the i-th training image is simply $\mathbf{g_i} = \mathbf{X_i} \cdot \mathbf{h}$. Where the dot product is defined in a complex space, and $\mathbf{X_i}$ is a diagonal matrix with $\mathbf{x_i}$ along its diagonal. The Average Correlation Energy is simply

$$NE_{ave} = \Sigma_i E_i = \Sigma_i \mathbf{g_i} \cdot \mathbf{g_i} = \Sigma_i \mathbf{h} \cdot \mathbf{D_i} \cdot \mathbf{h}$$

Where $\mathbf{D_i}$ is a diagonal matrix that contains the *power spectrum of the i-th image*. If the matrix $\mathbf{D}$ is defined as one that contains the *average power spectrum of the training set*, then:

$$E_{ave} = \Sigma_i \mathbf{h} \cdot \mathbf{D} \cdot \mathbf{h}$$

The MACE filter is chosen to minimize this quantity constrained by the requirement that the peak intensities for the training set, $\mathbf{X} \cdot \mathbf{h} = \mathbf{u}$. Where $\mathbf{X}$ is a matrix whose columns are the images inside the training set, and $\mathbf{u}$ is a predefined vector, usually with all components equal to one. By means of the Lagrange multipliers method it can be shown that the optimization process leads to:

$$\mathbf{h} = \mathbf{D}^{-1} \cdot \mathbf{X} \cdot \left(\mathbf{X} \cdot \mathbf{D}^{-1} \cdot \mathbf{X}\right)^{-1} \mathbf{u}$$

1.2.2 HBCOM

The Composite Filter, represented by HCOM is also based on a linear combination involving a subset from the subject's image database, such subset is known as the *Training set* or the *Reference set*. This type of linear filter is well consider for the intended application since it allows to merge several versions of reference images and it can also be useful to solve sensitivity issues thanks to rotations or scaling operations [1]. For the purpose of the project, a binarized version of this filter (call it HBCOM) is employed since it shows an advantage on its performance.
The way this filter can be constructed is by a simple standard linear combination, with the addition of a binarization [2]:

$$H_{BCOM} = sgn\left(\Re\left\{\Sigma_i a_i R_i\right\}\right)$$

Where $sgn(x)$ is the sign function which is defined as:

$$sgn(x) = \begin{cases} 1 & \text{if } x > 1 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 1 \end{cases}$$

Now, the weights a_i are specifically chosen to optimize some cost function which depends on the desire application, whereas the term R_i represents the Fourier spectrum of the i-th training image used to construct the filter [1].
As was mentioned before, the proper election of the linear combination coefficients is a crucial step to enhance the filter behaviour, ensuring its correct performance. Thus, one of these selection criteria is based on the **Peak-to-Correlation Energy** value or PCE, which is defined as the energy of the peak on the plane, normalized to the total energy of the plane.
Hence, when constructing the HBCOM filter by means of the PCE criteria is important to remark that apart from the reference set used on the linear combination, one must select another image from the database called *Filter image* whose purpose consists on serve as the sample from which the linear combination coefficients are to be found. Hence, based on the study done by de la Tocnaye, Quémener and Pétillot [2], some minor changes were carried out so the process to select the weights is done as follows:

First, compute individually all the Fourier transforms of the training set images, that is, calculate R_i for all values of i and also calculate the spectrum T for a suitable filter image, which for this specific process is going to be treated as a test image.
Then,perform the product $T^* \cdot R_i$ for all values of i and calculate the inverse Fourier transform in each case. This would be defined as a *Partial Correlation* or PC, since it represents the correlation between the target and a single reference. Therefore, if there are N samples on the training set, one gets virtually N different Partial Correlation Planes.
Now, define a total correlation plane as the inverse transform of the convolution between the target image and a spectrum compose of all the training samples. Thus, this plane is represented as:

$$C = FT^{-1}\left\{T^* \cdot \Sigma_i R_i\right\}$$

Note that C represents the total correlation plane for a case in which the convolution is computed between the test sample and a linear combination of the training images but with unit valued weights.
Now, since each partial correlation gives a different peak, a PCE value for each of them can be defined as the energy of the respective peak, normalized to the energy of the plane C defined above:

$$(PCE)_i = \frac{E_{peak,i}}{E_C}$$

Where $E_{peak,i}$ represents the peak energy on the i-th partial correlation and E_C is the energy of the total correlation plane described before. The respective energies are calculated as:

$$E_{peak,i} = \left|max((PC)_i)\right|^2 \quad E_C = \Sigma_{nm}\left|c(n,m)\right|^2$$

Where $max((PC)_i)$ represents the peak of the i-th partial correlation and $c(n,m)$ is the value assigned to the point n,m on the plane C and

correspondingly $|c(n,m)|^2$ is the energy of the point in question.
Finally, the weights are found as $a_i = [(PCE)_i]^{-\alpha}$. where α is is a constant value between 0.5 and 1.5. For this project the chosen value was α=0.8.
Therefore, the HBCOM filter is constructed by means of a sample of training images and a filter image which is treated as a target in order to find the optimal values for the linear combination coefficients. Thus, the filter is finally design and store for the respective subject under which it was made, this means that each individual is assigned their own filter, built following the previous steps.

1.3 Peak-to-sidelobe Ratio Normalization

When computing correlation planes it is readily seen that the peak intensity depends on the overall intensity of the query image and the training images used on the filter. Hence, a proper normalization metric is needed to address this problem. Those metrics are used as an indicator of the peak sharpness in the correlation output. Their value is of huge importance for devising a decision algorithm in the context of biometric verification. One of these metrics is the PSR, which measures the ratio of the peak height to its width. This metric is computed by a simple formula

$$PSR = \frac{peak - \mu}{\sigma}$$

In general, this metric is evaluated using convolution for the entire correlation plane. However, for simple applications, it is enough to select a patch in the correlation plane that contains the maximum value of the intensity, and compute mean and standard deviation over the patch.

1.4 ROC curves Analysis

The implementation of a simple classification algorithm for facial verification relies on a binary decision based on the normalization metric. Hence, by comparing the PSR value of a specific image with a suitable threshold, it is possible to decide if there is a match somewhere on the image plane. The assessment of this type of classifiers can be assessed in the **ROC space**, where the rate of accurately categorized images from a true-class dataset is compared to the rate of falsely identified matches from a false-class dataset. Subsequently, by considering the so called **confusion matrix** and the behavior of the classifier in this space it is possible to obtain not only the ideal threshold value, but also the expected error in the performance of the respective filter.
The calculation of a confusion matrix is based on the fact that when developing a classifier algorithm, some images will be falsely identified as matches, whereas some might possibly be wrongly classified as no-match. Therefore, the matrix is compose by four quantities meant to characterize the accuracy level of the biometric recognition application. Such values are defined as:

- True Positive Rate (TPR): Over a true-class dataset, it measures the ratio of accurately identified matches to the total number of true matches in the set. Ideally, the classifier should give a TPR = 1.
- False Negative Rate (FNR): Over a true-class dataset, it measures the ratio of wrongly rejected matches to the total number of true matches in the set. Ideally, the classifier should give a FNR = 0.
- False Positive Rate (FPR): Over a false-class dataset, it measures the ratio of wrongly identified matches to the total number of impostor images in the set. Ideally, the classifier should give a FPR = 0.
- True Negative Rate (TNR): Over a false-class dataset, it measures the ratio of accurately rejected matches to the total number of impostor images in the set. Ideally, the classifier should give a TNR = 1.

The upper quantities are dispose in the confusion matrix as it is shown in figure (3)

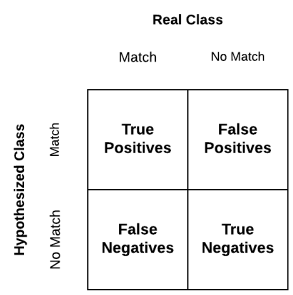


Figure 3: Confusion matrix structure.

Based on the previous descriptions, it is clear that an ideal behaviour would return an identity matrix. Moreover, each threshold has a particularly confusion matrix, in which ,for a very low value, virtually all images would be classified as matches and TPR = 1, FPR = 1. On the other hand, if it is too high, TPR = 0, FPR = 0. In both cases, the classifier is said to operate like a random classifier. Finally, a perfect performance is constrained by the actual metric computation and its ability to discriminate properly between true-class and false-class sets. Once the ideal threshold value is determined, the expected Equal Error Rate (EER) is the ordinate of the point in ROC space, as it is portrayed in figure (4).

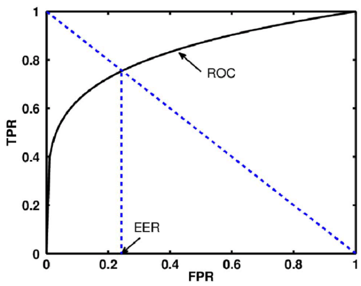


Figure 4: ROC space example [8].

The Receiver Operating Characteristic curve is obtained by plotting the TPR vs. FPR. The more accurate the classifier, the nearer would its ROC point be to the ideal (1,1). The random classifiers tend to locate the point near the principal diagonal of the plane (y = x). Usually, the curve is generated by varying the threshold metric value for the classifier. A good classifier must have a ROC curve above the principal diagonal of the space, otherwise such classifier is negated. Another measure of the accuracy, is the Area Under the Curve (AUC) of the ROC plot. Statistically, it measures the probability that more actual true matches are produced by the classifier than false positives. An AUC = 1 means that the classifier is completely accurate. An AOC = 0 means that the classifier is negated. And an AUC = 0.5 means that the classifier is likely to be random. Hence, the ROC curve and AUC can be used to assess the accuracy of the whole classifier algorithm.

2 Methodology

In order to achieve the objetives, 4 data sets corresponding to different subjects were acquired using conventional webcams with a resolution of 1280×720. An interface was developed to make snapshot alignment much easier. Additionally, illumination conditions were controlled by the use of common lamps, set up to achieve uniform illumination as closely as possible. A preprocessing protocol was established to reduce the noise in captured snapshots and obtain high discrimination rates. Once the data sets were captured, the filters were trained, using image registration, for each subject considered, and their performance assessed using ROC analysis.

2.1 Data acquisition

A special target was used to center the images in the data sets right from acquisition stage and facilitate separation of the actual region of interest (i.e. the face of the subject) from the background so that variations of facial expression with the face plane fixed and parallel to the camera were captured. The set up for data acquisition included 4 light sources with proper screening (i.e. a white paper sheet) coming from opposite directions: two lateral, one above and one bellow the face of the subject. The former light source improved significantly the quality of the acquired databases. For the project, 199 snapshots were captured per subject, and stored in PNG format.

2.2 Prepocessing techniques

The project focuses on performing biometric recognition on grayscale images. This implies using a proper conversion algorithm from RGB to gray. Image Preprocessing Toolbox from MATLAB was used in the preprocessing. Specifically, the function *rgb2gray*. This function converts and image from RBG to a grayscale by applying a weighted average over all red, green and blue components, reducing the image to a 2D matrix.

2.2.1 Wiener filter

By nature, MACE filters tend to whiten the image spectra which not only highlights biometric features, but also increases the noise level in the correlation plane. Hence, this type of filters is very sensitive to noise in the captured images. Since the imaging devices used were conventional webcams, the presence of random noise was inevitable. To correct that, a Wiener filter was used in the preprocessing stage. This filter is very effective in noise reduction and does not affect contrast dramatically. The application of de-noising before filter construction is improves noticeably the MACE performance. A sample of the influence of the Wiener filter is presented in figure (5). The correlation plane is less noisy and the peak has a smaller dispersion.

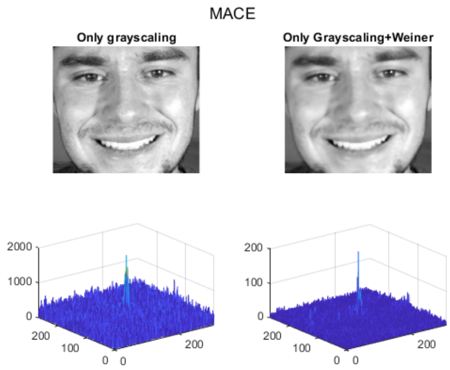


Figure 5: Correlation plane for MACE filter applied to grayscale images with (right) or without (left) Wiener filter.

For the particular case of HBCOM filter it is observed that binarization can reduce the noise in the correlation plane, and increase peak intensity. A sample is shown in figure (6). However, that preprocessing is not included because over the considered data sets it is not clear whether it can improve the discrimination power.

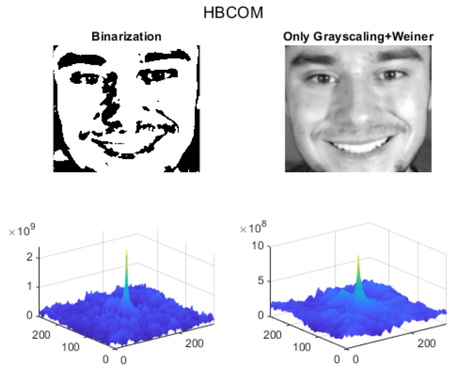
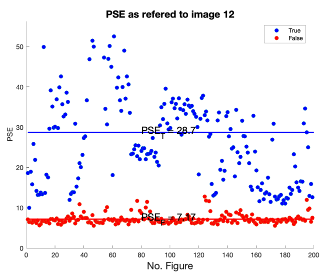


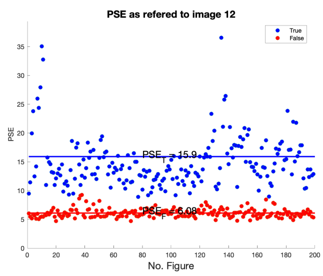
Figure 6: Correlation plane for HBCOM filter applied to grayscale images with (right) or without (left) Wiener filter.

2.2.2 Logarithmic normalization

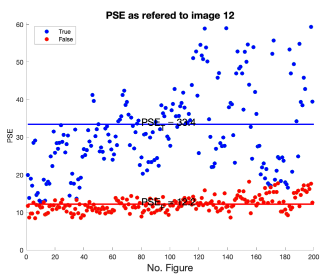
This preprocessing whitens the image plane and is a standard preprocessing technique used in most non-segmentation applications and it was the one applied on this work. A demonstration of the influence of the preprocessing in the discrimination capabilities of the correlation filters is shown in figure (7). The images show the comparative PSR distributions over true-class and false-class data sets for a filter devised using 2 samples from a database.



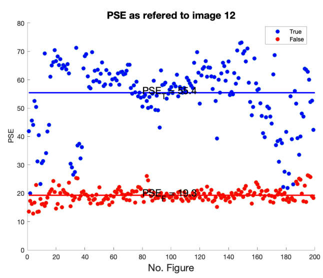
(a) Only RGB conversion.



(b) RGB + Log filter.



(c) RGB + Wiener filter



(d) RGB + Log + Wiener

Figure 7: PSR distributions over true and false class data sets for different preprocessing techniques combinations.

2.3 Filter construction and training

Image registration was automated inside the routines used. However, the discrimination capabilities of the filter depend on the set of training images used, their number and the template image used to base the filter construction process. This is illustrated as an example in the figure (8). Therefore, several training sets of different sizes were used to build filters and only those that produce filters with the highest discrimination between true class and false class were assessed in ROC space. The quantitative criterion was the difference between the mean PSR value of the correlation plane over true class sets and false class sets.

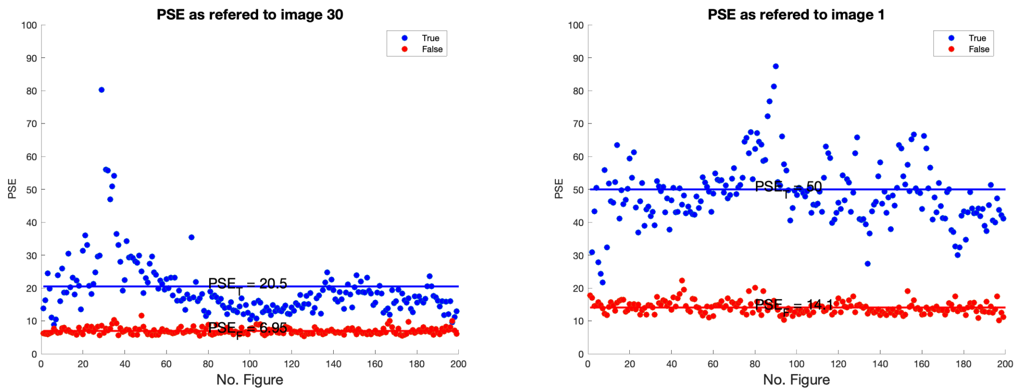


Figure 8: Illustrations of bad (left) and good (right) discrimination.

For the filters mentioned above, a classifier algorithm was associated. This algorithm simply thresholds the PSR value of the correlation output of a query image and makes a binary decision. By varying the threshold value, ROC plots were generated for filters associated to all subjects considered. Locating the threshold values that are nearest to the (0,1) corner in ROC space, the ideal parameters of the classifier were established. Furthermore, computation of confusion matrix is used to estimate the performance indicators of the biometric classification algorithms.

3 Simulations and analysis

In order to illustrate the capabilities of linear correlation filters for facial recognition, the captured datasets where used to develop a classifier based on facial features using MACE and HBCOM filters. By ROC analysis, the optimal threshold values for the classifiers were computed and the corresponding performance indicators were calculated. The results are presented below.

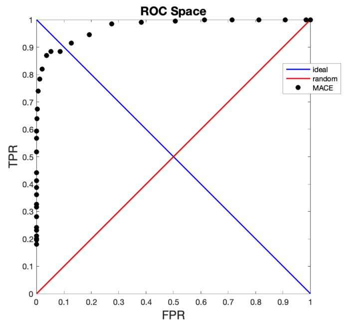
3.1 MACE

The corresponding simulations for this filter were carried out using the values shown in the table (1) for filter synthesis.

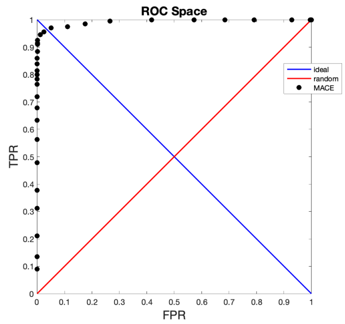
Subject	Ref. Imag.	Num. Images
A	2	2
D	12	2
G	102	2
N	39	2

Table 1: Parameters used in MACE synthesis for each subject's database.

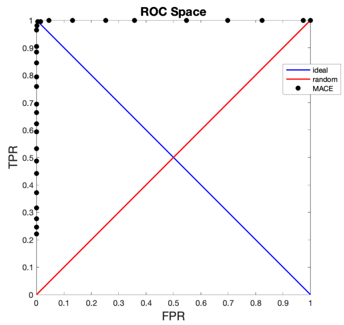
Afterwards, the ROC curves shown in figure (9) were obtained by following the process described in the methodology.



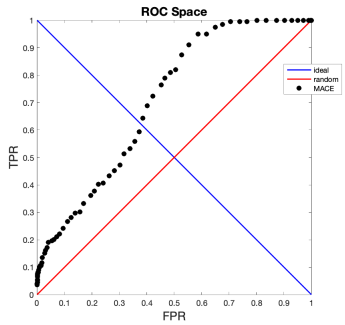
(a) Subject A.



(b) Subject D.



(c) Subject G



(d) Subject N

Figure 9: ROC curves from MACE filter.

From the plots in figure (9), the optimal threshold metric was computed, and also the performance indicators associated to the classifier algorithm.

Subject	Threshold	TPR	FPR	FNR	TNR
A	11.8633	0.8844	0.0854	0.1156	0.914
D	28.4959	0.9548	0.0241	0.0452	0.9759
G	16.4256	0.9950	0.0050	0.0050	0.9950
N	9.7330	0.6432	0.3849	0.3568	0.6151

Table 2: Performance indicators correlated to ROC curves from figure (9).

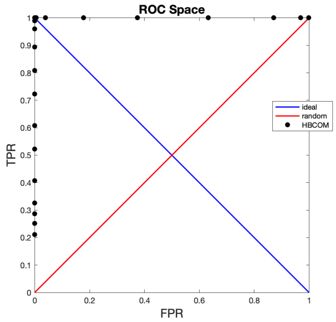
3.2 HBCOM

The corresponding simulations for this filter were carried out using the values shown in the table (3) for filter synthesis.

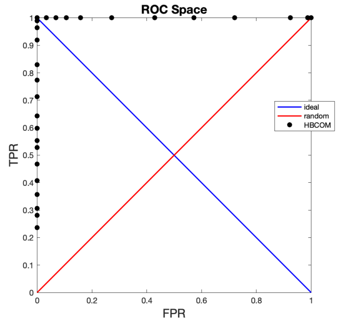
Subject	Ref. Imag.	Num. Images
A	46	4
D	1	4
G	40	4
N	165	4

Table 3: Parameters used in HBCOM synthesis for each subject's database.

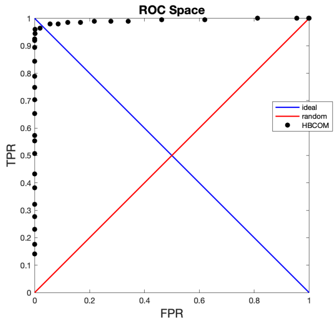
Afterwards, the ROC curves shown in figure (10) were obtained by following the process described in the methodology.



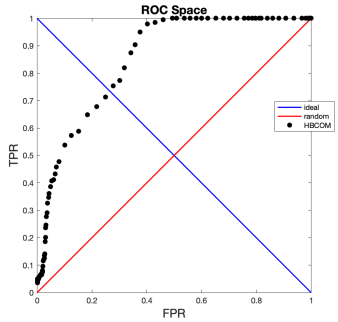
(a) Subject A.



(b) Subject D.



(c) Subject G



(d) Subject N

Figure 10: ROC curves from HBCOM filter.

From the plots in figure (10), the optimal threshold metric was computed, and also the performance indicators associated to the classifier algorithm.

Subject	Threshold	TPR	FPR	FNR	TNR
A	8.1815	1.000	0.000	0.0000	1.0000
D	8.5902	1.000	0.000	0.000	1.0000
G	7.7166	0.9648	0.0191	0.0352	0.9809
N	6.1542	0.7538	0.2774	0.2462	0.7226

Table 4: Performance indicators correlated to ROC curves from figure (10).

Based on these results, it is observed that the false positive rate is notably small for most of the subjects and likewise, true positive rate values are all above 80%, meaning that MACE and HBCOM filters present a sufficiently accurate discrimination capacity. As an additional note, is worth noticing that according to the confusion matrices, the best performance is obtained with the HBCOM filter.
Results for subject N are observed to be not as good as the others, which might be caused by errors in the data acquisition stage. Moreover, other 2 data sets were tested using the established protocol, and results turned out similar to subjects A, D and G, thus showing the high influence that a proper database has in non-segmentation biometric recognition. Those results are not included since publish permit from the subjects was not given.

4 Conclusions

It is found that both MACE and HBCOM filters show an acceptable discrimination considering that the false positive rate is significantly small in most cases and correspondingly, true positive rate values are all above 80%, which is the ideal range of accuracy set out as the ultimate goal. Furthermore, the HBCOM filter designed for this work generally exhibits a remarkable performance, as can be seen by the indicators. Consequently, the HBCOM filter is presented as the utmost result of the project in terms of reliability and efficiency.
On the other hand, results for subject N manifest some inconsistencies from the other subjects. This might be due to the fact that the images on the corresponding database could have included some background objects, since, subject's face has very different dimensions when compared to other subjects. Therefore, it is clear that proper database acquisition is fundamental
Overall, the performance of the correlation algorithms is impressive, and the results show

that for variations in facial expression, the technique produces EERs less than 10% for most datasets.
Finally, the preprocessing techniques tested in this work, proved to be valuable for increasing subject discrimination, specially for MACE filters. Further work might include variations of the facial plane in the datasets.

References

[1] Understanding Correlation Techniques for Face Recognition: From Basics to Applications, Face Recognition. Afalou et. al. Extracted from: Afalou

[2] Composite versus multichannel binary phase-only filtering. de la Tocnaye, Quemener & Petillot. Extracted from: de la Tocnaye

[3] A Technique for Optically Convolving Two Functions. C. S. Weaver and J. W. Goodman. Extracted from: Weaver

[4] Signal detection by complex spatial filtering. A. V. Lugt. Extracted from: Lugt

[5] Face Verification using Correlation Filters. Kumar et. Al. Extracted from: Kumar

[6] MACE Correlation Filter Algorithm for Face Verification ni Surveillance. Omidora et. Al. Extracted from: Omidora

[7] Academic Press Library in Signal Processing: Volume 4. Image, Video and Biomedical Signal Processing, and Hardware. Extracted from: Signal Processing

[8] New Perspectives in face Correlation Research: A Tutorial. Wang et. Al. Extracted from: Wang