A Face-based Authentication System Using Correlation Filters on Videos

José F. L. de Oliveira [†1], Eduardo A. B. da Silva [‡2], Manuel A. P. Cardoso [†3], Axel G. Hollanda [†4]

[†] *Instituto de Tecnologia José Rocha Sérgio Cardoso,*
Distrito Industrial, Manaus, AM
CEP: 69.075-210, Brasil.
[1] jleite@lps.ufrj.br
[3] manuel@internext.com.br
[4] aghi@lps.ufrj.br

[‡] *COPPE/PEE/LPS, Universidade Federal do Rio de Janeiro,*
Caixa Postal 68.504, Rio de Janeiro, RJ
CEP: 21.945-970, Brasil.
[2] eduardo@lps.ufrj.br

***Abstract*—This work is the result of the beginning of the development of a system for helping visually impaired people to recognize faces and objects. In order to make such a system, the problem of face recognition must be addressed and this is done by employing recognition algorithms, such as CFA – *Class-dependence Feature Analysis*, and a webcam. The objective of this work is to combine, improve, and develop algorithms for face detection and recognition so as to create a software-based system which is able to detect and recognize faces, previously enrolled in a database, obtained from images captured from a webcam. Specifically for the case of CFA, an algorithm for selecting the images that will compose the training set is proposed which reduces the training time by removing redundant images. The training time reduction is about 80%, without impacting identification performance, which is quite significant. Moreover, some techniques that employ the video sequence captured from the webcam in a very simple way are proposed for increasing verification reliability and stability.**

I. Introduction

The demand for algorithms capable of detecting and recognizing faces and/or objects has grown in recent years. Besides being used in applications like access control and conformity verification in production lines, such algorithms are beginning to be employed to help visually impaired people to better interact with their environment, giving them more independence and a feeling of safety. A device, which will arise from the development of this work, that is able to meet this latter class of applications is the LDV – *Lanterna para Deficientes Visuais* that means Flashlight for Visually Impaired People. In general, visually impaired people are able to identify people when they talk. With the LDV, the carrier of visual impairment, in addition to be able to become aware of the presence of people around him/her, can also "recognize" them without hearing their voices.

The objective of this work is to combine, improve and develop algorithms for face detection, such as Viola-Jones [1], [2], [3], [4], [5], and face recognition, such as CFA (*Class-dependence Feature Analysis*) [6] so as to devise a software-based system that is able to detect and recognize faces, previously enrolled in a database, from images captured by a webcam of low resolution, such as 320×240 pixels. Although there are several papers published on detection and recognition of faces and objects, there is a demand for works that explain in details the implementation of a system as the one that is going to be described here. The Viola-Jones algorithm [5] was chosen as a starting point for face detection because it is a very fast and efficient state of the art algorithm. For recognition, it was decided to develop an algorithm based on CFA. Of course, other techniques also known to be state of the art, such as EBGM (*Elastic Bunch Graph Matching*) [7], [8] or even KCFA (*Kernel Class-dependence Feature Analysis*) [9] were considered as a starting point. However, the lower computational complexity of CFA has favored its choice. The SIFT (*Scale-Invariant Feature Transform*) algorithm [10], [11], [12] is also appropriate for a system like this because of its generality, efficiency and relatively low complexity. A another system for recognition of faces and objects, based on SIFT, is also being developed by the authors but its description is out of the scope of this work.

Next, in section II, a background on correlation filters, and particularly on CFA, is presented. In section III, we describe the pre-processing of face images, that aims at reducing the effect of variations in illumination and pose. Section IV is dedicated to the presentation of some methods developed for reducing the CFA training time and increasing stability and reliability of identification. Results are presented in section V and conclusions in section VI.

II. Background

Correlation filters are basic tools for image matching in the frequency domain [13]. In methods employing them, normal variations in authentic training images can be accommodated by the design of an array in the frequency domain, called *correlation filter*, that captures the consistent part of the training images, deemphasizing frequencies that correspond to inconsistent parts [6]. Face recognition is carried out by the cross-correlation of an input image with a filter designed using the DFT (*Discrete Fourrier Transform*). Correlation filters also provide an incorporated shift-invariance. If the input image is displaced with respect to the ones of the training set, the peak output is shifted by the same amount. This shift can then be estimated by the correlation output and employed to align

[0]978-1-4244-4652-0/09/$25.00 ©2009 IEEE

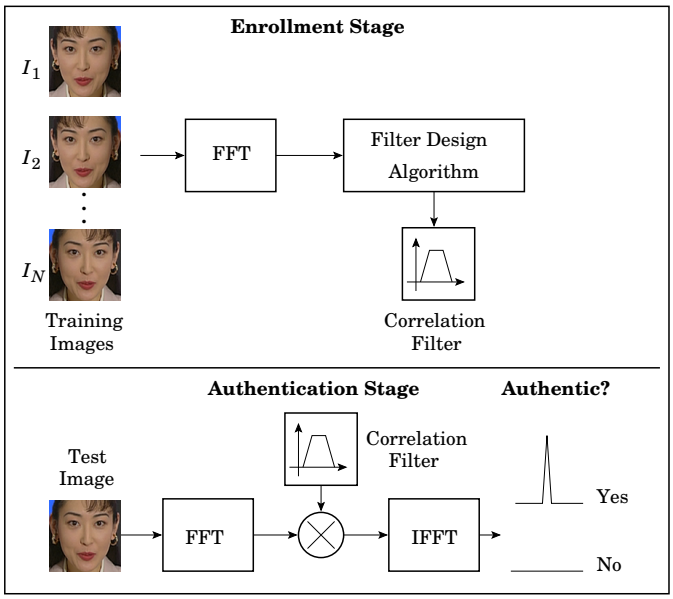

Fig. 1. Recognition of static image faces using correlation filters.

images. Another advantage of using correlation filters is that they provide closed form solutions which are computationally effective [6].

A. Face Recognition with Correlation Filters

The basic model of recognition of static face images by using correlation filters is shown in Figure 1 (adapted from [6]). There are two stages: one is for registration (enrollment) and the other is for recognition (authentication). In the registration stage, one or more images of a face of an individual are obtained. These images should reflect the expected variability in the face image due to rotation, scale changes, illumination changes, etc. The 2-D Fourier transform (2-D DFT) of these training images are then used by a correlation filter design algorithm to determine a single array in the frequency domain, the correlation filter itself. In the recognition stage, an image of a face is presented and its 2-D DFT is multiplied by the stored correlation filter. The 2-D inverse DFT of this product results in the correlated output. If the filter is well designed, a large peak in the correlation output should be observed if the face is recognized, and no discernible peak otherwise. The location of the peak indicates the position of the input image and thus automatically provides invariance to displacement; this makes it possible to avoid a centralization stage [6].

B. The Optimum Trade-off Filter

One way to design a correlation filter is to optimize one or more correlation criteria, under the restrictions of the correlation output c_j, which is the internal product of the training image and the filter to be determined

$$c_j = \mathbf{x}_j^T \cdot \mathbf{h}, \tag{1}$$

where $\mathbf{x}_j$ denotes the j-th training image and $\mathbf{h}$, the filter. The symbol "T" denotes the transpose of the complex conjugate. Typically, $c_j = 1$ for the training images of class "true" and $c_j = 0$ for those of class "false" [6].

Different criteria lead to filters with different properties. The filter MVSDF – *Minimum-Variance Synthetic Discriminant Function* filter – minimizes the variance of the noise of the correlation output $\mathbf{h}^T\mathbf{C}\mathbf{h}$, where $\mathbf{C}$ is a diagonal matrix whose elements C_{ii} represent the power spectral density of the noise in the frequency f_i. The MACE filter – *Minimum Average Correlation Energy filter* – minimizes the average energy of the correlation output $\mathbf{h}^T\mathbf{D}\mathbf{h}$. $\mathbf{D}$ is the mean of $\mathbf{D}_j$, a diagonal matrix whose elements D_{ii}^j represent the power spectrum of the j-th training image in the frequency f_i [6].

The MACE filter emphasizes high spatial frequencies to produce large correlation peaks, while the MVSDF filter, in general, removes the high frequencies for noise tolerance. Although it is desirable to meet both criteria, they cannot be minimized simultaneously. The optimal trade-off filter (OTF) is designed to balance these two criteria, by minimizing $\mathbf{h}^T\mathbf{T}\mathbf{h}$, where $\mathbf{T} = \alpha\mathbf{D} + \beta\mathbf{C}$, $\beta = \sqrt{1-\alpha^2}$ and $0 \leq \alpha \leq 1$. The OTF is then given by

$$\mathbf{h}_{\mathrm{OTF}} = \mathbf{T}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{T}^{-1}\mathbf{X})^{-1}\mathbf{c}, \tag{2}$$

where $\mathbf{X} = [\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}]$ is an $M \times N$ matrix and each $\mathbf{x}_j$ is the M-dimensional vector version corresponding to the 2-D Fourier transform of the j-th training image [6].

C. Derivation of the Optimum Trade-off Filter

The solution given by equation 2 is obtained finding the filter $\mathbf{f}$ which minimizes the cost function $\mathbf{\Phi}(\mathbf{f}) = \mathbf{e}^T\mathbf{e}$, where $\mathbf{e} = \mathbf{c} - \mathbf{Y}^T\mathbf{f}$. Then,

$$\begin{aligned}\mathbf{\Phi}(\mathbf{f}) &= \left(\mathbf{c}^T - \mathbf{f}^T\mathbf{Y}\right)\left(\mathbf{c} - \mathbf{Y}^T\mathbf{f}\right) \\ &= \mathbf{c}^T\mathbf{c} - \mathbf{c}^T\mathbf{Y}^T\mathbf{f} - \mathbf{f}^T\mathbf{Y}\mathbf{c} + \mathbf{f}^T\mathbf{Y}\mathbf{Y}^T\mathbf{f}\end{aligned} \tag{3}$$

and, thus,

$$\frac{\partial\mathbf{\Phi}}{\partial\mathbf{f}} = -2\mathbf{c}^T\mathbf{Y}^T + 2\mathbf{f}^T\mathbf{Y}\mathbf{Y}^T, \tag{4}$$

since $\partial\mathbf{A}\mathbf{x}/\partial\mathbf{x} = \partial\mathbf{x}^T\mathbf{A}^T/\partial\mathbf{x} = \mathbf{A}$ and $\partial\mathbf{x}^T\mathbf{A}\mathbf{x}/\partial\mathbf{x} = \mathbf{x}^T\left(\mathbf{A}^T + \mathbf{A}\right)$. Then, $\partial\mathbf{\Phi}/\partial\mathbf{f} = 0$ implies $\mathbf{Y}^T\mathbf{f} = \mathbf{c}$. If the columns of $\mathbf{Y}$ are linearly independent when $M \geq N$, the matrix $\mathbf{Y}^T\mathbf{Y}$ of dimension $N \times N$ is inversible. That way,

$$\mathbf{Y}^T\mathbf{f} = \mathbf{Y}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{c} = \mathbf{c}, \tag{5}$$

and this implies $\mathbf{f} = \mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{c}$.

Equation 2 is obtained if $\mathbf{Y} = \mathbf{T}^{-1/2}\mathbf{X}$ is replaced in equation 5, since $\mathbf{T}^{-1/2}\mathbf{f} = \mathbf{h}$. Note that, if the diagonal matrix $\mathbf{T}$ is inversible, so $\mathbf{X}^T\mathbf{T}^{-1}\mathbf{X}$ is too, when $M \geq N$. The matrix $\mathbf{T}$ is used as an approximation of the whitening of the training set, that comprises the columns of matrix $\mathbf{X}$.

D. Class-dependence Feature Analysis

Class-dependence feature analysis trains a bank of correlation filters based on data from a generic training set, where there are multiple genuine images of each class. The set of correlation filters is then used in validation experiments to extract discriminant class dependent features for recognition.

However, when this algorithm is used in a face recognition device, such as the LDV, the training set would be initially empty, and later, as new individuals (classes) were registered, the number of classes would increase progressively. It happens that, in order to the CFA algorithm perform satisfactorily, the number of initial classes should be large enough so as to allow

a proper training [6], [9] of the correlation filters. The solution proposed in this paper to solve this problem is to divide the training set into two parts. One is composed by individuals who should actually be recognized, named *primary individuals* or *primary classes*. The other is composed by individuals who serve only to provide the initial number of classes necessary for adequate training of the correlation filters, referred to as *secondary individuals* or *secondary classes*. If there is a significant correlation peak on a secondary individual, the system simply returns a negative identification.

A problem that arises with this kind of solution is that an individual in a primary class can be similar to one in a secondary class. In this case, there could be high peaks in both classes, creating difficulties in the recognition. This problem is partially solved with the algorithm developed in subsection IV-A.

Another problem that needs to be solved is the interpretation of the correlation vector produced by the CFA algorithm (the inner product of the DFT of input image with the correlation filters associated to each individual in the training set). One has to determine when a correlation peak is significant or not. A solution to this problem is discussed in section IV-B.

III. Pre-processing of Face Images

Generally, one of the first problems to be solved for the implementation of a face recognition system with a webcam results from the following fact: face recognition algorithms such as CFA need the input images to have their position and illumination normalized. This aims at minimizing the negative influence of their variations during the recognition procedure. In what follows we present an illumination normalization method (proposed in [14]) and a position normalization method proposed in this work

A. Position Normalization

Typically, the normalization of the position of a face uses the coordinates of eyes because those points are easy to locate. Once their coordinates are known, a translation followed by a scale transformation and rotation is sufficient to normalize the face's position. For a face recognition system that acts on still images, eye location with human intervention is not a serious problem. In the case of dynamical identification of face images generated by a webcam, it is not practical to use such a procedure. To circumvent this problem, the Viola-Jones detector is used to obtain the coordinates automatically. Although the detector determines the coordinates of the eyes with an error, as long as it makes an error of the same nature for the images in the training database, it is compensated along the recognition process.

B. Illumination Normalization

The illumination normalization method described below was proposed in [14] and consists of three main steps: gamma correction, difference of Gaussian filtering – DoG *Filtering* – and contrast equalization.

1) Gamma Correction: Given a user defined parameter $\gamma \in [0, 1]$, the gamma correction is a a non linear transformation of the gray levels. It replaces the value $I(x, y)$ by $I^{\gamma}(x, y)$ for $\gamma \neq 0$ or by $\log(I(x, y))$ when $\gamma = 0$. It has the effect of expanding the local dynamical range in the dark or shade image regions and shrinking it in white and very illuminated regions. An exponent in the range $[0, 0.5]$ represents a good trade-off. The value $\gamma = 0.2$, suggested in [14], has been adopted as a standard value.

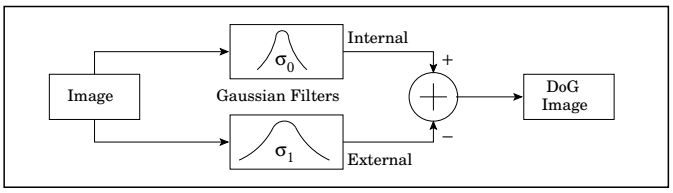

Fig. 2. Difference of Gaussian Filtering – DoG Filter.

2) Difference of Gaussian Filtering: The gamma correction does not remove the influence of all intensity gradients such as the ones generated by shading effects. Shading induced by the surface structure is potentially a useful recognition cue but is predominantly visual information from low spatial frequency, that is difficult to separate from the effects caused by gradients of illumination. High-pass filtering removes both useful and incidental information, thus simplifying the problem of recognition, and in many cases, increasing the total system performance. Similarly, by removing the highest spatial frequencies, aliasing and noise are both reduced. In practice, this must be done without affecting too much the part of the signal which recognition is based on.

DoG filtering, see Figure 2, is a convenient way to obtain the resulting band-pass behavior. Fine spatial details are critically important for recognition, so the internal Gaussian filter is typically narrow with $\sigma_0 \leq 1$ pixel, while the external filter can have σ_1 ranging from 2 to 4 or more pixels, depending on the spatial frequency at which the information of low frequency becomes more misleading than informative. For data sets with large variation of illumination, [14] recommends $\sigma_1 \approx 2$. For less extreme variations in illumination, a variance value of up to 4 pixels can be used.

3) Contrast Equalization: This step globally rescales the image intensities to normalize a robust measure of the whole contrast range. It is important to use a robust estimator because the input images typically contain a small mix of extreme values. These are produced by the most illuminated regions of the image, garbage on the edges of the image and dark regions, such as the nostrils. The process of equalization is done in two steps as shown in equations 6 and 7

$$\begin{aligned} I_0(x, y) &= |I(x, y)|^a \\ I_1(x, y) &= \frac{I(x, y)}{\bar{I}_0^{1/a}} \end{aligned} \tag{6}$$

$$\begin{aligned} I_2(x, y) &= [\min\{\tau, |I_1(x, y)|\}]^a \\ I_3(x, y) &= \frac{I_1(x, y)}{\bar{I}_2^{1/a}}, \end{aligned} \tag{7}$$

where $a < 1$ is a strongly compressive exponent that reduces the influence of the high values, and τ is a threshold used to truncate the high values after the first stage of normalization. $\bar{I}_0$ and $\bar{I}_2$ are the mean values of $I_0(x, y)$ and $I_2(x, y)$, respectively. As standard values, $a = 0.1$ and $\tau = 10$ are used.

The resulting image is now properly scaled but still contains extreme values. To reduce their influence, it is finally applied

a hyperbolic tangent for compressing them, which limits $I_3(x, y)$ to values in the range $(-\tau, \tau)$.

$$I_4(x, y) = \tau \tanh\left[\frac{I_3(x, y)}{\tau}\right], \tag{8}$$

IV. A Face-based Authentication System

In this section, the algorithms and techniques developed to design a face-based authentication system using CFA are described. Firstly, it is presented the algorithm for selection of training images, which operates effectively in reducing the training time of CFA. Then, the heuristic created to decide whether a correlation peak is significant or not is described. This heuristic can set the level of reliability of the identification and how fast it is executed. Finally, it is presented the method developed to treat the vector of correlation peaks which contributes to identification stabilization.

A. Selection of Training Images

To calculate the filters used in the CFA, N_C individuals (classes) with N_{T_c} training images each (samples) must be provided, resulting in a total of $N_I = \sum_{c=0}^{N_C-1} N_{T_c}$ images. In general, to obtain correlation filters with good recognition capabilities, the training images should be as representative as possible of each class. Also, it is important that the images corresponding to each class are, if possible, in non-intersecting regions of $\mathbb{R}^N$. As, in practice, this may not occur, as in the case of twins, what can be done is to try to reduce these areas of intersection. Another problem that may occur is to have very similar images within the same class. Although this does not represent a problem for the recognition performance, it causes an unnecessary increase of the training time. Here we propose a simple method to alleviate these problems. It is based on the angle between two images $\mathbf{U}$ and $\mathbf{V}$ defined by

$$\cos(\theta) = \frac{\mathbf{U} \cdot \mathbf{V}}{|\mathbf{U}||\mathbf{V}|}. \tag{9}$$

If the images have zero mean and unit variance, then

$$\cos(\theta) = \frac{\mathbf{U} \cdot \mathbf{V}}{HW - 1} \tag{10}$$

for a non-biased variance estimative, where H is the height and W is the width of $\mathbf{U}$ and $\mathbf{V}$. Let θ_s be the image separation threshold. Then,

$$\begin{cases} \text{if } \theta \leq \theta_s & \Rightarrow \quad \text{images are similar} \\ \text{if } \theta > \theta_s & \Rightarrow \quad \text{images are non-similar.} \end{cases} \tag{11}$$

Note that, for zero mean and unit variance images, the classes are distributed on the surface of a hypersphere of radius $\sqrt{HW - 1}$.

Employing this criterion, while the list of candidate training images is scanned, an image is actually included in the training set if it is *not* similar to any other that has been previously selected for training. Figure 3 illustrates in two dimensions the performance of the proposed method. Note that if an image in a primary class is similar to one in a secondary class (see section II-D), it will be removed. If a secondary class is empty it is simply removed from training. If a primary class gets empty a solution is to reduce the separation angle θ_s and apply the algorithm again.

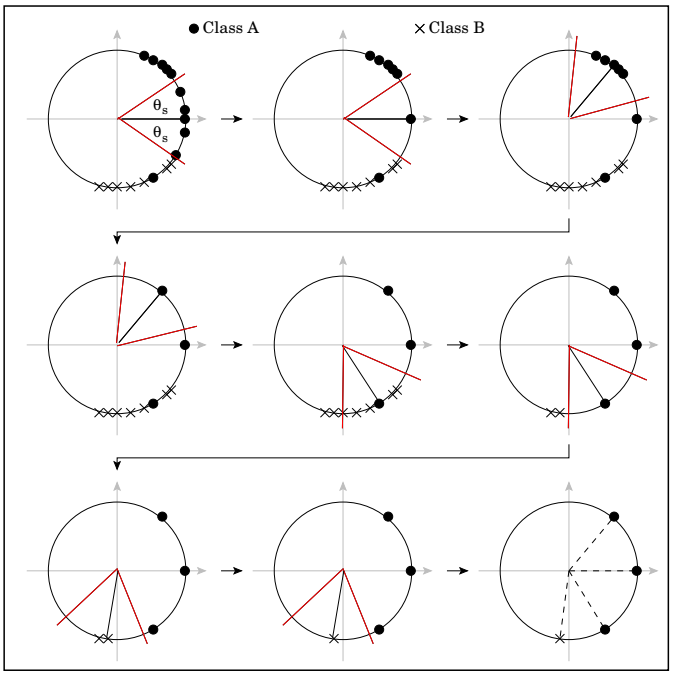

Fig. 3. Reducing the redundance in the list of training images.

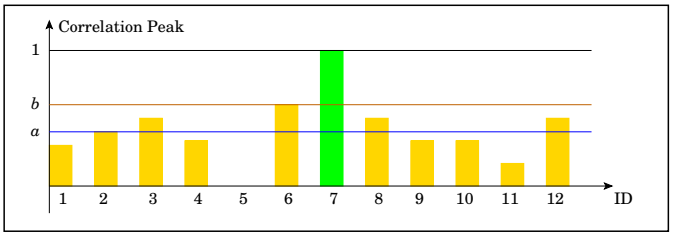

Fig. 4. Heuristic employed in the face-based recognition system using CFA.

B. Interpretation of the Vector of Correlation Peaks

What is considered a significant correlation peak must be defined. Since the correlation vector obtained by the CFA algorithm is real, the highest peak value, normalized to 1, corresponds in principle to the face identified. As there will always be a maximum, even if the input face has is not from a person registered to be recognized, care must be taken in order to have a stable and reliable positive identification. To this end, while the person to be identified is filmed, a counter C counts the number of frames for which the global maximum, m_g (ID 7, figure 4) falls on the same individual. If the individual changes or if the second maximum m_s (ID 6, in Figure 4) is greater than b, the counter is reset. It is considered that an individual has changed when face detection is interrupted, such as when an individual leaves the front of the camera. Then, if $m_s \leq a$, it is considered that an individual was identified with a single frame. If $a < m_s \leq b$ and $C \geq N_F$ it is considered that an individual was identified after N_F or more consecutive frames. That is, when $m_s \leq a$, a single frame is sufficient to consider the identification as being reliable. When $a < m_s \leq b$, at least N_F consecutive frames are required to have a reliable identification. The values adopted for a, b and N_F will be discussed in section V.

C. Stabilization of the Vector of Correlation Peaks

During the development of the face-based authentication system using CFA, it has been noted that second maximum values above the threshold b occurred randomly among the classes, while the global maximum tended to remain stable. As a consequence, the counter C was frequently reset, making it difficult to identify an individual. To reduce the effects of this problem we have developed the method described below.

Let $\mathbf{c}(n) = [c_0(n), c_1(n), \ldots, c_j(n), \ldots, c_{N_C-1}(n)]$ be the vector of correlation peaks for the n-th frame, where N_C is the number of classes. Since the camera provides a video sequence, the sequence of vectors of correlation peaks obtained for the registered individuals (see figure 4) can be averaged in order to obtain a better recognition. We perform a weighted average so that the more recent peaks have larger weights than the earlier ones. In order to do so, the m-th more recent correlation peak is multiplied by a forgetting factor λ^m, $\lambda < 1$. The smaller λ is, the faster the influence of a correlation peak occurred in the past is forgotten. The average vector of correlation peaks $\bar{\mathbf{c}}(n)$ becomes, after normalization,

$$\bar{\mathbf{c}}(n) = \frac{\sum_{m=0}^{N_W-1} \lambda^m \mathbf{c}(n-m)}{\max_j \left\{ \sum_{m=0}^{N_W-1} \lambda^m \mathbf{c}(n-m) \right\}}, \qquad (12)$$

where N_W is the number of frames inside the window used to compute $\bar{\mathbf{c}}(n)$. When $n < m$, $\mathbf{c}(n-m) = 0$. The values adopted for N_W and λ will be discussed in section V.

V. Results

The results described in this section have been obtained using a computer powered by an Intel® Pentium® Core 2 Duo 3 GHz processor, with 2 MB of cache and 2 GB of RAM, connected to a webcam Logitech® QuickCam Chat and running a Linux operating system.

A. Selection of Training Images

Initially, we have assessed the algorithm developed in section IV-A, which aims to select only a representative subset of training images used to obtain the bank of correlation filters. The assessment has been made in terms of its impact on training time and recognition capacity. In table I, the relative pre-processing time, T_p, and the relative training time, T_t, are shown as a function of the separation angle θ_s. The training set contains 46 classes: 42 of them are secondary classes (see section II-D) with 60 images each; 3 are primary classes with 600 images each; 1 is a primary class with 360 images. These result in a total of 4,680 images of dimensions 80×92. Each image is normalized in both position and illumination. The individuals in the primary classes were also used to compose the secondary classes. This had the purpose of testing the functionality of the algorithm of section IV-A in performing the elimination of the effects of interference that a secondary class may have on an individual of a primary class.

When $\theta_s = 0^\circ$ the pre-processing time includes only image position and illumination normalization of the training set. This is so because, in principle, due to capture noise introduced by the camera and by the variation of the environment illumination, there would not be two identical images in this set, Thus, for $\theta_s = 0^\circ$ there is no need to use the algorithm for selection of training images. The minimum value of $\theta_s > 0^\circ$, adopted in table I for the effective application of the algorithm of selection was chosen based on the calculation of the minimum angle between classes. Secondary classes that have corresponding primary classes were excluded from this calculation.

TABLE I
Training time T_t and pre-processing time T_p as a function of the separation angle θ_s used in the optimization of the list of training images.

$\theta_s(^\circ)$	$\frac{T_p}{T_p^*}$	$\frac{T_t}{T_t^*}$	$\frac{T_p+T_t}{(T_p+T_t)^*}$	$\frac{T_t}{T_p}$	$N_I(\%)$
0	0.14	1.00	1.00	39.19	100
45	1.00	0.21	0.39	1.13	57
50	0.81	0.13	0.27	0.84	45
55	0.62	0.07	0.18	0.61	35
60	0.45	0.03	0.11	0.56	24
65	0.31	0.01	0.07	0.20	13

The symbol "*" denotes the maximum value the variable has taken.

In spite of T_p for $\theta_s > 0^\circ$ be greater than the one for $\theta_s = 0^\circ$, the removal of redundant images dramatically reduces the training time T_t, causing a significant reduction in the total time ($T_p + T_t$). When there are individuals that appear in both primary and secondary classes, there is a significant improvement in their recognition ability until $\theta_s = 55^\circ$. Furthermore, when individuals are all distinct, this process causes no deterioration in recognition ability. In both cases, the total time ($T_p + T_t$) for the calculation of the filter $\mathbf{h}$ is significantly reduced, by about 80% for $\theta_s = 55^\circ$, as shown in table I. For this training set, $\theta_s = 55^\circ$ was the largest separation angle that did not produce observable degradation of recognition.

B. Stabilization and Interpretation of The Correlation Vector Peaks

The values a, b and N_F, which determine whether a correlation peak is significant or not, were determined experimentally. Before that, a range of values for λ and N_W were determined so as to the stabilize the correlation peaks. It was observed that, for favorable illumination conditions, the values $\lambda = 0.8$ and $N_W = 15$ were a good trade-off between stabilization and recognition speed. On the other hand, the values $k = 1$ and $N_W = 50$ were more appropriate for environments with unfavorable illumination. We have concluded that, in general, values of λ in the range $[0.8, 1]$ and N_W in the range between 15 and 50 were effective in stabilizing the correlation peaks.

With the correlation peaks stabilized, the next step was the experimental determination of the a, b and N_F. Initially, the threshold-b rule is disabled and several individuals, registered or not, were used to test recognition. The threshold a was initially set at 0.5 and had to be reduced to about 0.18 so that there were no false positives when using only a single sequence frame. For this a value, the registered individuals have positive identification only if there are excellent illumination conditions. The next step was to determine values for b and N_F. The system could process online the video generated by the camera at a rate of 15 frames per second including pre-processing, detection and recognition. Thus, it was decided to use $N_F = 15$, that is, at least one second would be necessary to make a positive identification of the individual

when $a < m_s \leq b$. Then, the threshold b was initially set at 1 and was gradually reduced until no false positives were observed. The value found was $b = 0.31$.

C. Performance Evaluation

In general, performance tests involving face recognition algorithms follow the procedures described in [15] to obtain the corresponding TAR (*True Accept Rate*) and FAR (*False Accept Rate*). In order to provide advanced statistical analysis, it has been estimated that a database with about 50,000 images was necessary. This database was in fact created by taking photos of 200 individuals per week for a year, generating the FRGC Ver2.0 (*Face Recognition Grand Challenge Ver2.0*) database [15].

A reliable measurement of the TAR and FAR of the face recognition system proposed in this work should follow a similar procedure, but using video sequences instead. A database with 50,000 video sequences of resolution 320×240 pixels (25 times less pixels than the smaller image size in FRGC Ver2.0), with 10 s each in duration and 15 frames per second, would have about 2 Tbytes. While storage space for 2 Tbytes is not a big problem nowadays, creating a database of this magnitude requires a considerable amount of time. Therefore, since a database like this is not available (at least to the authors' knowledge), an alternative strategy to evaluate the TAR and FAR was adopted, although less accurate.

Initially, the adjustment of the system parameters has been carried out as described in sections V-A and V-B. Then, 20 individuals outside the training set and that did not take part in the parameter adjustment procedure were used to both "validate" the configuration and estimate the FAR. All of the 20 individuals are researchers in image and signal processing. They were informed in advance how the system operated to recognize faces. A graphic with the correlation peaks was shown in real time for each individual that, with the help of the first author, tried to cause the occurrence of false positives, by changing pose, facial expression and/or changing the distance from the camera. All subjects had at least five minutes to try to accomplish this task.

Only one individual has caused the occurrence of false positives, but only when it moved away from the camera, which has no automatic focus control. In this case, besides the images analyzed by the algorithm being outside focus, position normalization also contributed to a smoothing of the image through the scale transformation. Details were lost and the chances of the system making errors increased. Besides that, three individuals registered to be recognized were instructed to proceed in the same way as those outside the training set so as to induce identification mistakes, but all of them were identified correctly.

D. Final Considerations

The face detection and recognition system has been designed to deal primarily with frontal faces and under controlled illumination conditions; however, the tests have shown that even with changes in face position (10 to 20° in horizontal and vertical directions) and small variations in the illumination intensity, recognition is still possible. How much the position of the faces may vary depends on the illumination conditions as well as on the diversity of images of faces that were provided for training.

VI. Conclusion

Some methods have been developed in this work that allowed the CFA face recognition algorithm to be more efficiently used in a video-based authentication system that acquires the faces from a webcam. The developed methods provided:

a. the use of Viola-Jones face detector for obtaining rapidly and automatically the approximate coordinates of the eyes, allowing automatic position normalization (section III-A);
b. reduction of the redundancy of the training set, leading to a significant reduction in training time of about 80%, without impacting the recognition performance (section IV-A);
c. the determination of the correlation peak significance, increasing the reliability of identification (section IV-B);
d. stabilization of the vector of correlation peaks, allowing that the identification was more stable too (section IV-C).

The experimental results also showed that the system remains functional for non-frontal face views and with small variations in illumination intensity.

References

[1] P. VIOLA and M. J. JONES, "Robust real-time object detection," Cambridge Research Laboratory/Compaq Computer Corporation, Cambridge, Massachusetts 02142, USA, Technical Report Series CRL 2001/01, February 2001.

[2] ——, "Robust real-time object detection," Vancouver, Canada, July 13 2001.

[3] ——, "Fast and robust classification using asymmetric adaboost and a detector cascade," in *Advances in Neural Information Processing System 14*. MIT Press, 2001, pp. 1311–1318.

[4] ——, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, Kauai, HI, USA, December 8-14 2001, pp. I.511–I.518.

[5] ——, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[6] C. XIE, M. SAVVIDES, and B. V. K. V. KUMAR, "Redundant class-dependence feature analysis based on correlation filters using frgc 2.0 data," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 3, San Diego, California, USA, June 20-25 2005, pp. 153–158 (6).

[7] L. WISKOTT, J.-M. FELLOUS, N. KRÜGER, and C. von der MALSBURG, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, July 1997.

[8] ——, *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, isbn: 0-8493-2055-0 ed. L.C. Jain et al. (Editores) CRC Press, 1999.

[9] R. ABIANTUM, M. SAVVIDES, and B. V. K. V. KUMAR, "How low can you go? low resolution face recognition study using kernel correlation feature analysis on the frgcv2 dataset," in *IEEE Biometrics Symposium*, Baltimore, Maryland, USA, September 19-21 2006.

[10] D. G. LOWE, "Object recognition from local scale-invariant features," vol. 2, Kerkyra, Grécia, September 20-27 1999, pp. 1150–1157.

[11] ——, "Local feature view clustering for 3d object recognition," vol. 2, Kauai, Hawaii, December 2001, pp. I–682–I–688.

[12] ——, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[13] M. SAVVIDES, B. V. K. V. KUMAR, and P. KHOSLA, "Face verification using correlation filters," in *Proceedings of Third IEEE Automatic Identification Advanced Technologies*, Terrytown, New York, USA, March 14-15 2002, pp. 56–61.

[14] X. TAN and B. TRIGGS, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proceedings of the 2007 Analysis and Modeling of Faces and Gestures*, Rio de Janeiro, Brasil, October 20 2007, pp. 168–182.

[15] P. J. PHILIPS, P. J. FLYNN, T. SCRUGGS, and K. W. BOWYER, "Overview of the face recognition grand challenge," vol. 1, San Diego, California, USA, June 20-25 2005, pp. 947–954.