Improved Face × Non-Face Discrimination Using Fourier Descriptors Through Feature Selection

TEÓFILO EMÍDIO DE CAMPOS
ROGÉRIO SCHIMIDT FERIS
ROBERTO MARCONDES CESAR JUNIOR

Instituto de Matemática e Estatística
Universidade de São Paulo (IME/USP),
Rua do Matão, 1010, PO-Box 66.281,
05315-970 São Paulo, SP, Brasil
{teo, rferis, cesar}@ime.usp.br

Abstract. This work presents a new method to discriminate face from non-face images using Fourier descriptors. The first step of our approach consists in applying a horizontal edge detection filter in the input image, followed by the extraction of a 1D signal from the computed edge map. Then we calculate the Fourier descriptors from this signal and classify the image using statistical classifiers. In order to improve our results, we applied a feature selection algorithm. Preliminary performance assessment results have shown that this approach is superior than traditional transform based methods. Besides, these results showed that our method might be used to develop a fast face detection system.

1 Introduction

Human face detection is an active research area in the computer vision community, with important applications in automatic face recognition, visual surveillance and man-machine interfaces. Furthermore, a very rich literature on biological face recognition can be found due to the importance of this task for humans and other animals.

Among a large number of proposed techniques, we can cite the work of Rowley, Baluja and Kanade [1], which uses a neural network to perform frontal face detection. Sung and Poggio [2] also presented a technique for locating vertical frontal views of human faces, with an example-based learning approach. These approaches provide good results, but, like the majority of the methods of the literature, their computational cost is very large, being not suitable for real-time face detection.

Recently, the work of Wu, Chen and Yachida [3] described an effective method to detect faces in color images based on the fuzzy theory. Another color-based approach to detect and to track faces is described in the work of Feris, Campos and Cesar [11]. These color-based approaches are efficient with respect to real-time processing. But they are not robust to presence of lots of skin color objects near the face in the image. Further problems of this approach include the fact that color images involves larger data sets than gray scale images and that the most of security systems use gray scale cameras instead of color cameras.

In this paper, we propose a new method to discriminate face from non-face images using Fourier descriptors. The preliminary obtained results are very promising, showing that we can use our approach to perform a task such as face detection, by sliding a window at different image locations and scales, or by detecting skin-color blobs, i.e., face candidates.

In order to discriminate face from non-face images, we have used Fourier descriptors, an effective technique to perform shape analysis, which has been successfully applied in a large number of different problems (see the paper of Zahn and Roskies [4] for further information). Basically, given an image to be classified as face or non-face, we first apply a horizontal edge detector, followed

0-7695-0878-2/00 $10.00 © 2000 IEEE

by the extraction of a 1D signal from the computed edge map. We then calculate Fourier descriptors from this signal and classify the image using the Fourier descriptors to form a feature vector. Preliminary results showed that our approach might be very useful to develop a fast and robust face detection system, mainly because only a few coefficients have been used in the classification. To improve our results, we applied a feature selection algorithm based on a search method proposed recently, and using, as the criterion function, the performance of the minimum distance to the prototype classifier.

The remainder of this paper is organized as follows. Section 2 describes the classification method and the assessing experiments, while Section 3 presents the experimental results. Finally, in Section 4 follows a conclusion, where we address further research directions.

2 The face/non-face discrimination method

The experiments carried out in order to investigate the proposed technique are summarized schematically in figure 1. The basic idea of the proposed approach is to extract a 1D signal from the horizontal edge map and to characterize the face patterns by a set of Fourier descriptors extracted from this signal. The obtained features have been evaluated by training a statistical classifier and classifying a test set in order to obtain the classification accuracy. In order to improve the system's performance, it was applied a feature selection algorithm over the Fourier coefficients. Each step is described below.

2.1 Face database formation

A set of 219 almost upright face images and of 219 non-face images has been used. The images do not present necessarily the same illumination conditions because different face databases that can be found in the Internet have been merged.

Some images created for our database have also being included. The latter images have been acquired using a FujifilmDX-7 digital camera, with resolution of 640 X 480 pixels. The images have been hand-cropped so that only eyes nose and mouth are taken into account. The face images database and the pictures considered have the property that all the obtained face crops are larger than 128 x 128 pixels. An example is shown in figure 2.

2.2 Non-face database formation

The non-face image database has been created by randomly searching for images in the WWW. A vision research software environment, called Synergos, has been implemented (see [5]), which includes computational tools for searching and storing WWW images.

Synergos includes a web-worm-like robot that makes the access to HTML documents, i.e. that implements the http protocol in order to exchange data with WWW servers. The resources that Synergos may extract by parsing a HTML document include other HTML documents, downloadable files, images, ftp links, emails, newsgroups and so on. An important feature of this approach is that the nature of the obtained images vary a lot, for the search engine navigates through many different web pages depending only on the hypertext structure (i.e. the URL's of the visited page).

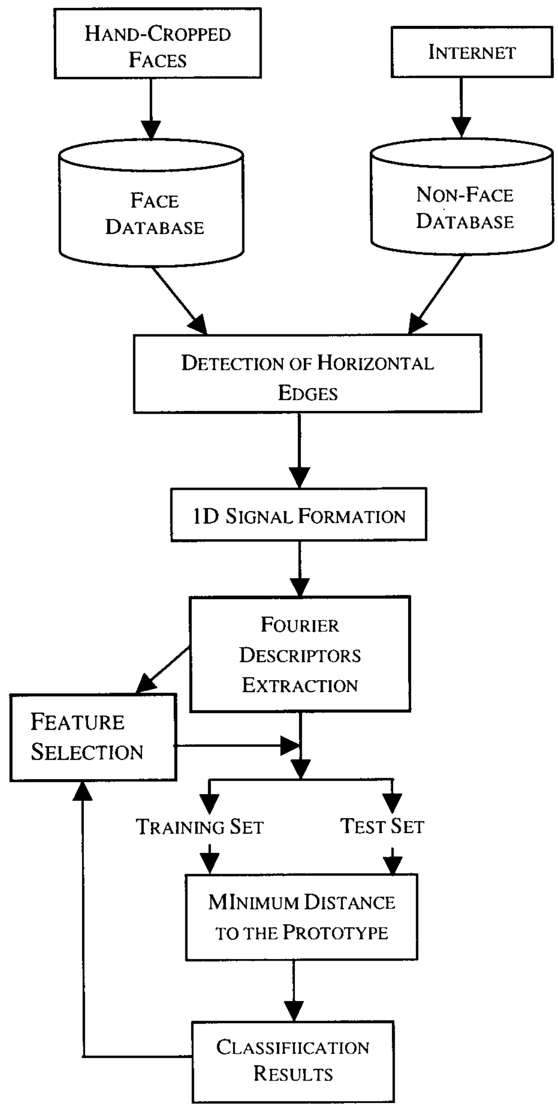
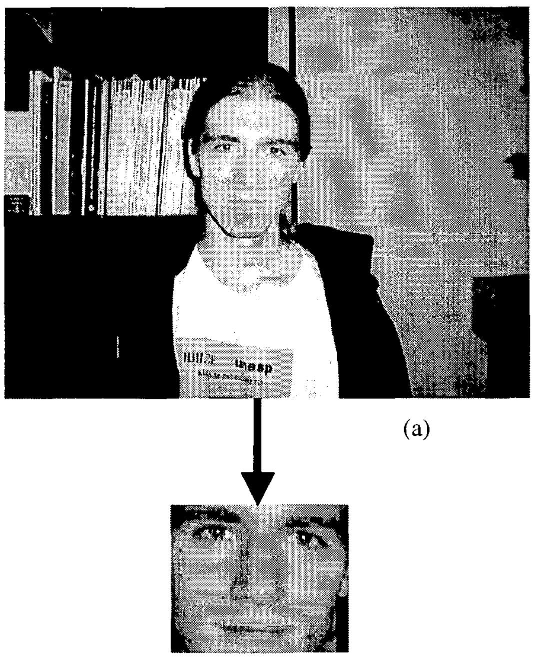


Figure 1 General overview of the experiments architecture.

Once that a WWW page is visited, the Synergos gets the images therein. In order to normalize the database, the following procedure is adopted: Synergos selects only images larger than 256 X 256. Then, it chooses randomly a block of size 256 X 256 from the selected image and this block is stored in the database. In order to verify that no face image has been selected, a human operator examines the database (though images that present faces, among other features, are allowed).



(a)

(b)

Figure 2 Input image obtained with a digital camera (a) and its respective cropped face image containing only eyes, nose and mouth (b).

2.3 Size normalization

All images from both databases are normalized with respect to size, so that all images are 128 X 128 before the subsequent steps. There are many reasons for performing size normalization. First, for classification purposes, between faces and non-faces, it is important to normalize size because comparisons between features of the feature space are actually made.

Furthermore, the search for a minimum size for feature extraction is also very important in order to save computational demand. In fact, because the proposed approach rely on the discrete Fourier transform, it is also important to have signals of suitable size, so that Fast Fourier Transform (FFT) algorithms can be adopted.

Finally, normalizing the image size provides a homogenization for using the same parameter values among the different imaging algorithms for all input images. Size normalization is a step commonly used in many different face detection procedures, such as those described by Sung and Poggio [2].

2.4 Detecting of horizontal edges

Although some previous works on face detection and recognition have already suggested the use of image edges (like Moghaddam and Pentland [6], and Kondo and Yan [7]), the experiment reported in this work shows that very good correct discrimination rates between faces and non-faces can be achieved using only the horizontal edges of face images.

The horizontal edges preserve mainly the information of eyes, eyebrows, mouth and the region around the nostrils. There are two basic approaches for exploring the information associated to horizontal edges: using an edge emphasized image (e.g. differentiated); and using an edge map (hence, a binary image), which can be obtained, for instance, by thresholding or by searching for local maxima of the edge emphasized image.

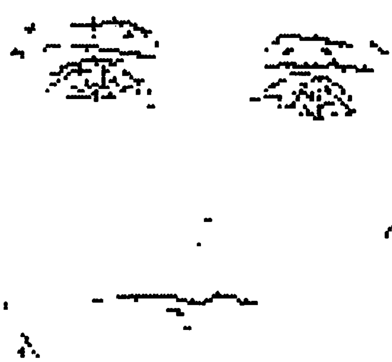


Figure 3 Horizontal edge map.

The traditional methods for edge enhancement and detection include the gradient based and Laplacian based techniques, such as Roberts, Sobel and the Laplacian-of-Gaussian. In the experiment reported in this paper, a simple Sobel mask for detecting horizontal edges has been used, providing very good results.

Figure 3 presents the edge map of figure 2 (b) using Sobel. It is important to note that the visual information regarding the face structure is preserved.

2.5 1D signal formation

The main idea behind the method proposed here is to characterize the edge patterns formed from the edge detection step. An approach borrowed from 2D shape analysis techniques is based on extracting some 1D signature from the pattern followed by a linear transform of this signal, say the Fourier or the wavelet transform (see Antoine et. al. for further details [8]).

There are many different ways of defining the signal, such as Scholl diagrams, shape matrices, the pattern image itself and so on. Because the experiment discussed in this work assumes vertical and frontal faces, an invertible representation of the edge map can be obtained taking the first column of the image, followed by the second, the third, and so on.

Let $h(p, q)$ be the m X n edge map, with $h(p, q) = 1$ if the pixel (p, q) is an edge point and 0 otherwise, where $p = 0, 1, ..., (m-1)$ and $q = 0, 1, ..., (n-1)$. Therefore, the characterizing signal is defined as:

$$g(p + mq) = h(p, q)$$

Clearly, $g(i)$, $i = 0, 1, ..., (mn)$ is a discrete time binary signal, or an impulse train. Figure 4 shows the signal $g(i)$ obtained from the edge map of Figure 3.

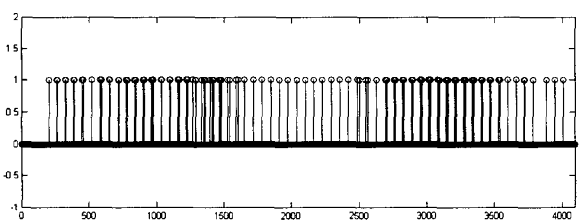


Figure 4 1D signal (impulse train) representing the edge map of figure 3.

2.6 Fourier descriptors

The problem that must be tackled now is how to characterize signals $g(i)$ obtained from the edge maps in order to discriminate faces from non-faces. Clearly, a simple point-to-point comparison between the signals would not lead to good results for two very similar edge maps could have very few impulses coinciding exactly at the same time locations.

A preferable approach is to define and obtain the Fourier descriptors of $g(i)$ [4] and to use them as a feature vector together with a statistical pattern classifier. There are many different Fourier descriptors (FD's) in the literature, and they may differ a lot depending on the problem in hand. In this work, we define the FD's of $g(i)$ as:

$$G(s) = \log\left(\left|\sum_{i=1}^{mn} g(i)\, e^{-j2\pi \frac{si}{mn}}\right|\right), s \in S, S \subset F$$

where F is the set of all the FD's, S is a subset of F. The FD's are based on the modulus of the complex coefficients of the Fourier transform of $g(i)$. Adopting the modulus equips the FD's with the nice property of invariance with respect to time shifting of $g(i)$ (which only affects the phase of the Fourier transform). The *log* function is important because the coefficient amplitudes may vary a lot, which could be a problem when using them together in feature vector comparisons by the statistical classifier. The log function attenuates such large coefficient variations. Finally, it is important to note that the FD's defined above do not include the 0-th DC component, which only depends on the number of edge points in the image, despite their spatial organization.

2.7 Feature selection

The above mentioned feature extraction process produces a large number of features (FD's) that can be used for classification. Nevertheless if we use a large number of FD's, the performance of the statistical classifier will decrease with respect to execution time and to recognition rate. The execution time increases with the number of features because of the measurement cost. The recognition rate can decreases because of redundant features and of the fact that small number of features can alleviate the course of dimensionality when the number of training samples is limited. On the other hand, a reduction in the number of features may lead to a loss in the discrimination power and thereby lower the accuracy of the recognition system. Jain, Duin and Mao's paper [14] made a good review about these facts.

Once that real-time face detection is aimed, it is desirable to define a feature space that can lead to good classification rate with a simple and fast classifier. Therefore, it is necessary to keep our pattern representation (Fourier space) as small as possible. In this context, the problem is how to select the best FD's set.

In order to investigate this question, we made some tests with the first 30 FD's obtained through our technique. Basically, given the desirable feature set size d, these tests are divided into 3 categories, that will be described in the next subsections.

2.7.1 First d FD's

This is the most straightforward ways to select features in a Fourier space for image analysis. It is based on the fact that the most of the signals' energy from real images are concentrated on the low frequencies, so we can discard high frequencies, which are redundant for doing classification.

2.7.2 Largest d FD's

This approach is often adopted in the context of recognition using wavelets. It is based on the idea that the most important coefficients of a linear transform are the largest coefficients. Therefore, the FD's are chosen by doing a threshold. In case of wavelet research, an analogous technique is known as wavelet shrinkage.

In our case, we selected these features by first determining the mean vector of all the samples of the face class and taking the index of the d largest elements of this vector. Alternative voting schemes can be adopted if the features are not so homogeneous for specific classes.

2.7.3 Feature selection using search methods

Automatic feature selection is an optimization technique that consists on, given a set of m features, select a subset of size d that leads to the maximization of some criterion function. In the case of this work, we used the performance of a classifier as criterion function, as explained in section 3.

There are many different feature selection methods. A well know algorithm is the Branch and Bound (BB), that is an optimal method for monotonic feature sets. Nevertheless its computational cost is prohibitive for large feature sets: in the worst case, its time complexity is exponential on the dimension of the feature space.

In 1997, Jain and Zongker published a review paper with a taxonomy and a comparative study about the feature selection methods [10]. According to them, probably the most effective feature selection techniques was the sequential floating search methods (SF), proposed by Pudil et al. in 1994 [12].

Basically, in the case of forward search (SFFS), the algorithm starts with a null feature set and, for each step, the best feature that satisfies some criterion function is included with the current feature set. The algorithm also verifies if the criterion can be improved if some feature is excluded. Therefore, the SFFS proceeds dynamically increasing and decreasing the number of features until the desired d is reached. The backward search works analogously, but it starts with the full feature set. The time complexity of these methods is linear.

In 1999, Somol et al. proposed some improvements on these algorithms [13] and created the adaptive floating search methods (ASF). The difference between this new approach and the previous is on the number of features that can be inserted or excluded from the feature set. The new algorithms can determine dynamically this number, while the earlier tests one feature per step. The advantage of ASF is that its results are closer to the optimum solution than the results of SF methods. However, for large values of d, ASF is very slow.

Both methods do not have the restriction that the feature space has to be monotonic, like the BB method.

For these reasons, we choose to do tests with automatic feature selection algorithms based on SF and ASF. Our tests comparing the performance of feature spaces generated by these four aforementioned methods will be explained in section 3. We omit further detail about the SF and ASF algorithms because of paper length restrictions. The reader is referred to [12 and 13].

2.8 Statistical classifier

A statistical classifier together with the feature vector formed by the FD's can be used in order to discriminate between face and non-face images. We implemented the "minimum distance to the prototype" statistical classifier in which one prototype per class is defined, where each prototype is the average of all training samples. The advantage of this classifier is its computationally efficiency (its time complexity is $O(n)$, for n being the number of features). This classifier is further described in the Duda and Hart book [9].

Besides the simplicity of this classifier, it is possible to reach a good correct classification performance. This fact is illustrated in the next section.

3 Experimental results

We have used the "minimum distance to the prototype" statistical classifier to obtain our results. The database contains 219 face images and 219 non-face images, all of them have resolution normalized to 128 x 128 pixels.

Our tests are summarized in the next tables. The results considered the recognition rate on classifying faces and non-faces in percentages. We performed tests varying the size of the desired feature set d, and using the four methods to determine the feature set.

Let Λ be the learning set of our classifier, T the test set and U be the full set of the database. We did tests with $\Lambda \cap T = U$ and others tests with $|\Lambda| = 66{,}67\%$ of $|U|$ and $|T|$

= 33,33% of |U| and Λ∩T={ }. The elements of the sets Λ and T was chosen randomly.

In order to apply the feature selection (SF and ASF methods), we used as criterion function the classification rate of the statistical classifier based on minimum distance to the prototype. This criterion function perform the classification on sets that Λ∩T={ }. Therefore, the SF and ASF methods work to maximize the classification rate using learning and testing sets with no intersection.

Table 1 shows the some of the obtained results for a face database |U| = 219 and non-face database |U| = 219. Figure 5 resumes of the results of the minimum distance to the prototype classifier when tested in images that was not in the learning set (Λ∩T={ }) for all the values of d tested.

4 Concluding remarks

We have proposed a new method to discriminate face from non-face images using Fourier descriptors that show very promising results.

The best results were obtained through ASF methods, being superior to the traditional methods to perform classification using Fourier coefficients. The most traditional approach for selecting the FD's coefficients is by using the first d's and it presented the worst results (except for $d \geq 15$, in this case, the "largest d" is the worst).

Another important fact is that, for all tested methods, the best results was obtained for $d<20$. This result confirms that increasing the number of features does not guarantee that the performance of the classifier also increases.

The main difference between the performance of the SF and ASF was 4,3478%, for d=13 and d=15. But in the worst case (concerning execution time), SF spent 2 seconds do reach the best subset (for d=25), while, in the same computer, ASF spent 4 hours, 22 minutes and 10 seconds (for d=15). These facts show that if the priority is to obtain fast results on the feature selection, it is better to choose the SF algorithms. It is important to see that even if SF is not the best feature selection algorithm, its results are better than the results of "First d" and "Largest d", for all cases.

An important fact is that 30 is the total number of features available and, for this reason, all the methods leads to the same result when d=30.

It is important to emphasize three facts about the face database:

- these tests were done using hand cropped face images, without precise registration;
- the faces were not exactly upright;
- the considered database is composed by several face databases obtained with different equipment and under different illumination conditions.

Table 1 Classification rate tables (%).

d=3	Λ∩T=U	Λ∩T={}
First *d*	66,1836	71,0145
Largest *d*	66,1836	71,0145
SF	69,5652	75,3623
ASF	69,5652	75,3623

d=9	Λ∩T=U	Λ∩T={}
First *d*	74,8792	73,9130
Largest *d*	74,8792	73,9130
SF	78,7440	76,8116
ASF	81,1594	78,2609

d=15	Λ∩T=U	Λ∩T={}
First *d*	79,2271	59,4203
Largest *d*	79,7101	57,9710
SF	85,5072	75,3623
ASF	86,4734	79,7101

D=21	Λ∩T=U	Λ∩T={}
First *d*	82,6087	50,7246
Largest *d*	81,6425	49,2754
SF	85,5072	75,3623
ASF	85,5072	75,3623

d=27	Λ∩T=U	Λ∩T={}
First *d*	85,5072	56,5217
Largest *d*	84,5411	55,0725
SF	84,5411	65,2174
ASF	84,0580	66,6667

As expected, these database characteristics are very prejudicial to the classifier's performance. Nevertheless, they show the robustness of our system with natural conditions for face detection.

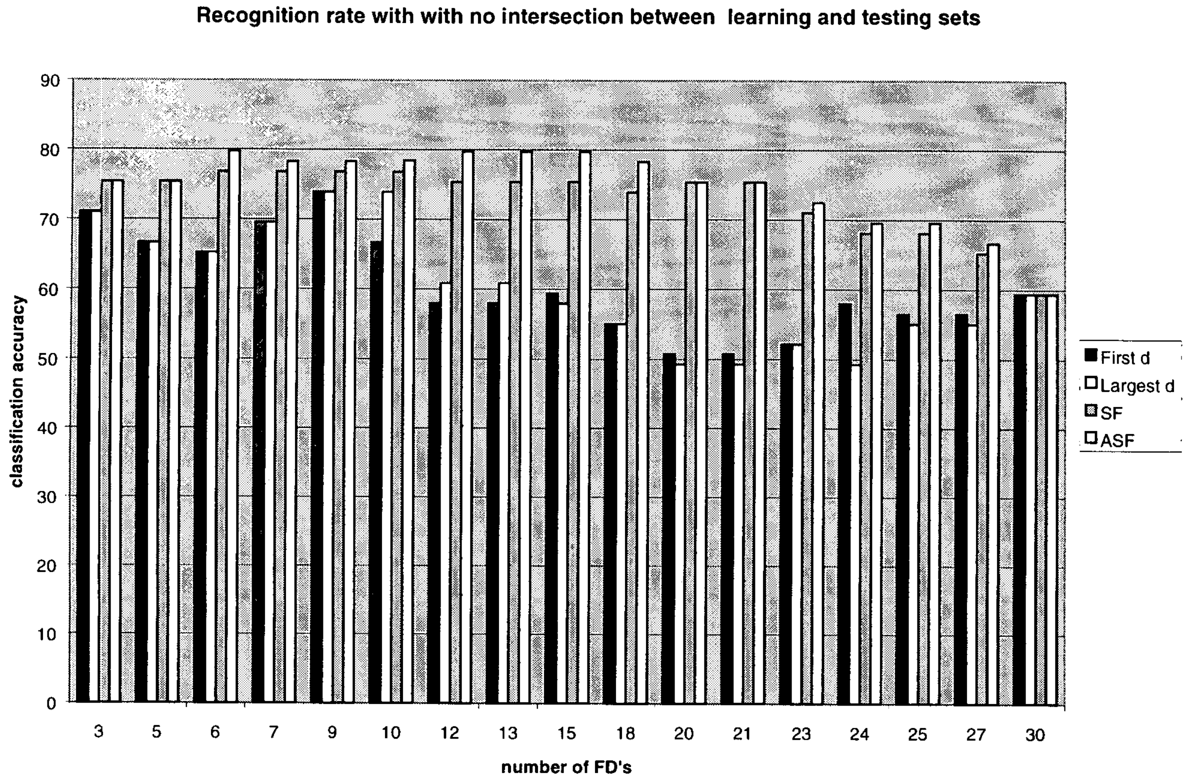


Figure 5 Results of the classifier for $\Lambda \cap T = \{\}$.

The best result obtained was the recognition rate of 86,4734% (d=15 and $\Lambda \cap T = U$ with ASF), which is a very promising result. Nevertheless, the main contribution of this work is the speed of the method. For vectors with only 3 dimensions, it is possible to detect faces with an accuracy of 74,3243% minimum distance to the prototype classifier.

As a future work, we plan to apply other feature extraction methods, other criterion function for feature selection and to perform these tests using other classifiers.

5 Acknowledgements

Teófilo E. Campos is grateful to FAPESP for the financial support (99/01488-8), Rogério S. Feris is grateful to FAPESP (99/01487-1) and Roberto M. Cesar J. is grateful to FAPESP (98/07722-0) as well as to CNPq (300722/98-2).

We are grateful to P. Somol, P. Pudil, J Novovicová and P. Paclík for providing some source codes.

6 References

[1] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(1) (1998), 23--38.

[2] K.K. Sung and T. Poggio, "Example-based learning for view-based human face detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(1) (1998) 39--55.

[3] H. Wu, Q. Chen, and M. Yachida, "Face Detection From Color Images Using a Fuzzy Pattern Matching Method", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21(6) (1999) 557--563.

[4] C.T. Zahn & R.Z. Roskies, "Fourier descriptors for plane closed curves", *IEEE Trans. on Computers* C21 (1972), 269--281.

[5] O. M. Bruno, R.M. Cesar Jr., L.A. Consularo, and L. da F. Costa, "Automatic Feature Selection for Biological Shape Classification in ΣYNERGOS", *Proc. Brazilian Conference on Computer Graphics, Image Processing*

and Vision, (SIBGRAPI-98, Rio de Janeiro - RJ, Out 1998), IEEE Computer Society Press (1998), 363--370.

[6] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) (1997), 696--710.

[7] T. Kondo, H. Yan, "Automatic human face detection and recognition under non-uniform illumination", *Pattern Recognition* 32(10) (1999), 1707--1718.

[8] J.-P. Antoine, D. Barache, R.M. Cesar Jr., L. da Fontoura Costa, "Shape characterization with the wavelet transform", *Signal Processing*, 62(3) (1997), 265--290.

[9] R.O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, NY, 1973.

[10] A. Jain and D. Zongker, "Feature selection - evaluation, application, and small sample performance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2) (1997), 153--158.

[11] R. S. Feris, T. E. Campos and R. M. Cesar J., "Detection and tracking of facial features in video sequences", *Lecture Notes in Artificial Intelligence,* vol. 1793 (2000) 127--135, Springer-Verlag.

[12] P. Pudil, J Novovicová and J. Kittler, "Floating search methods in feature selection", *Pattern Recognition Letters* 15 (1994), 1119--1125.

[13] P. Somol, P. Pudil, J Novovicová and P. Paclík, "Adaptive floating search methods in feature selection", *Pattern Recognition Letters* 20 (1999), 1157--1163.

[14] A. Jain, R. P. W. Duin and J. Mao, "Statistical pattern recognition: a review", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 22, no. 1 (2000), 4--37.