International Journal of Pattern Recognition and Artificial Intelligence
Vol. 16, No. 3 (2002) 349–359

A SUPERVISED MODIFICATION OF THE HAUSDORFF DISTANCE FOR VISUAL SHAPE CLASSIFICATION

ALBERT PUJOL* and JUAN JOSE VILLANUEVA

Centre de Visio per Computador and Universitat Autonoma de Barcelona, 08193 Bellaterra, Barcelona, Spain
**albert@cvc.uab.es*

JOSE LUIS ALBA

Communications Technology Department, University of Vigo, 36200 Vigo, Spain

Hausdorff distance is a deformation tolerant measure between two sets of points. The main advantage of this measure is that it does not need an explicit correspondence between the points of the two sets. This paper presents the application to automatic face recognition of a novel supervised Hausdorff-based measure. This measure is designed to minimize the distance between sets of the same class (subject) and at the same time maximize the distance of sets between different classes.

Keywords: Supervised Hausdorff measures; face recognition; face shape descriptor; Hausdorff distance; shape similarity.

1. Introduction

Automatic Face Recognition (AFR) has been a successful field of research mostly during the past two decades. This is evident when considering the continuous publication of reviews and surveys, from the earliest by Samal and Iyengar,[16] to the latest by Grudin,[6] passing through the works of Valentin *et al.*,[18] and Chellapa *et al.*[3] Besides the impressive development in the research field, some face recognition problems still require further development,[14,15] this is the case for problems of recognizing face images conveying changes in illumination, facial expression, and changes due to the time delay between the acquisition of the reference and tested face images. These changes are due to two sources of variability, the first source produces a change in the intensities received by the image sensors, meanwhile the second one produces a rearrangement of the intensities across the sensing area.

Currently, some of the most successful AFR systems try to solve these two problems by applying statistical pattern recognition methods directly to gray level face images. Some of these approaches allow, given an adequate training set, to define features and similarity measures where the expected distance between two images of the same subject (within class distances) is minimized, and the expected distance between images of different subjects (between class distances) is maximized. Examples of these methods are the "Fisherfaces"[1] and the "dual eigenspaces".[12]

The difference from the above approaches, is that the work presented in this paper is based on the use of an illumination invariant feature descriptor, and a measure of the distortion or rearrangement of the positions where the features have been detected.

The work presented in this paper uses as image feature extractor the MLSEC operator,[10] obtaining image valleys which are, by definition, invariant to illumination changes. The relevance of valleys as face shape descriptor has been pointed out in some cognitive science works.[13] The response of a valley detector (second row at Fig. 2) depicts the face in a similar way to that of a human drawing, showing the position, and extent of the main facial features. In addition, our method uses as a similarity measure a novel modification of the Hausdorff distance. Hausdorff-based distances[4,7] are flexible matching techniques that allow to measure distortion (displacement or position rearrangement of the image features) rather than changes in image intensities. Unlike the previously defined modifications of the Hausdorff distance, our proposal uses the distortions observed in a training set to derive a more adequate similarity measure for face image recognition.

The contents of this paper are distributed as follows. In Sec. 2 a general framework is introduced in order to review some of the Hausdorff-based distances that can be found in the literature. In Sec. 3 the application of these measures to image face recognition is described. The framework introduced in Sec. 2 will allow us to describe, in Sec. 4, the proposed supervised Hausdorff measure. In Sec. 5 the experimental results obtained using the new shape similarity measure are compared, in terms of recognition ratios, with those obtained with previously defined Hausdorff-based measures. Finally Sec. 6 summarizes some conclusions and provides some areas of interest that may be worthy of further research.

2. Hausdorff Measures

Hausdorff distance is a measure of similarity between two sets of points belonging to the same metric space. The main advantage of this measure is that an explicit correspondence between points is not required. This requirement is overcome through an implicit nearest-neighbor correspondence between the points of the sets. Formally, given two sets of points A, and B, the Hausdorff distance $H(A, B)$ between both sets is defined as:

$$H(A, B) = \max(h(A, B), h(B, A)) \tag{1}$$

$$h(A, B) = \max_{a \in A}\left(\min_{b \in B}(d(a, b))\right) \tag{2}$$

$h(A, B)$ being the direct Hausdorff distance from the set A to the set B, and $d(a, b)$ the distance, usually Euclidean, between the points a and b. It has to be noted that the direct Hausdorff distance is not in general symmetric, and that some of the Hausdorff distance modifications that can be found in the literature are not metrics. For this reason we will refer to them as Hausdorff-based measures. In order to analyze the different Hausdorff-based measures in a common framework, the definition of the Hausdorff distance can be generalized, rewriting the directed and undirected Hausdorff distance as:

$$H(A,B) = f^s(h(A,B), h(B,A)) \tag{3}$$

$$h(A,B) = f^i_{a\in A}\left(f^m\left(\min_{b\in B} d(a,b)\right)\right) \tag{4}$$

We will call "*symmetrizer*", "*integral*", and "*best-match*" function to f^s, $f^i_{a\in A}$, and f^m, respectively. Using this framework, Hausdorff distance would be obtained when:

$$f^s(x,y) = \max(x,y) \tag{5}$$

$$f^i_{a\in A}(f(a)) = \max_{a\in A}(f(a)) \tag{6}$$

f^m being the identity function. Different symmetrizer and integral functions have been proposed in the literature, among others the $K^{x\text{th}}_{a\in A}$ rank-order distances (giving rise to the generalized Hausdorff distance[7]). Dubuisson and Jain,[4] tested the robustness to noise of the Hausdorff distance and some of their possible modifications in an image edge matching experiment. In this experiment each image was modeled as the set of edge pixel locations. Different *integral* (minimum, maximum, $K^{25\text{th}}$, $K^{50\text{th}}$, $K^{75\text{th}}$, and average) and *symmetrizer* functions (minimum, maximum, average, and weighted average) were considered. After testing all possible combinations of the considered *integral* and *symmetrizer* functions, the authors concluded that the most robust measure to synthetic noise was the Modified Hausdorff Distance (MHD). The MHD is obtained when average and maximum are considered respectively as *integral* and *symmetrizer* functions. Besides the above mentioned robustness of MHD to synthetical noise we conducted a first experiment comparing the robustness of those measures in a set of real face images (see Sec. 5), from these preliminary experiments we concluded that better recognition results were obtained when the average was used both as *symmetrizer* and *integral* functions. We have denoted the recognition results obtained using this measure as AHD (standing for average Hausdorff distance) in Table 1. In order to overcome the problem of outliers, some methods define a tolerance measure so that a matching between two points will be considered only if the distance between them is smaller than a certain amount. This is the case of the Hausdorff fraction.[8] The Hausdorff fraction between two sets A and B, for a fixed Neighborhood of size d, measures the portion of the elements of the set A, that are on the inside of a

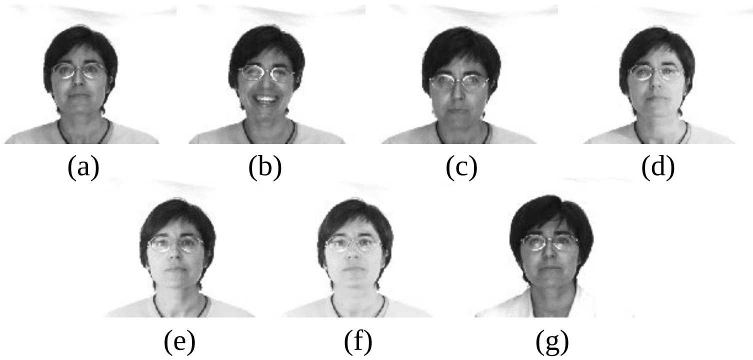


Fig. 1. Examples of the face images used in the experiments.

neighborhood of radius d at some point of the set B. In this case,

$$f^i_{a\in A}(f(a)) = \frac{1}{|A|}\sum_{a\in A} f(a), \text{ and } f^m(x) = \begin{cases} 1 \text{ if } x < d \\ 0 \text{ otherwise} \end{cases} \tag{7}$$

The main drawback of this technique is that an exhaustive search of the neighborhood size d, must be done in order to apply this measure.

3. Using Hausdorff Measures as Image Similarity Criteria

We are concerned with image face recognition. Before introducing the supervised Hausdorff distance (SHD), it is convenient to explain, how Hausdorff measures (HM) are applied to face images and how they can be computed in an efficient manner. Hausdorff based distances have been previously used for face detection[5] and recognition.[17] Both works used as image feature descriptor the response of an edge detector.

Examples of the face images and acquisition conditions used in the experiments reported later are shown in Fig. 1. These images are a subset of the AR Face database.[11] Images are normalized in size and orientation using the eye positions. The normalized images are then cropped obtaining images of 78 × 68 pixels like those shown in the first row in Fig. 2.

Image Ridges and Valleys are then extracted using the *multilocal level set extrinsic curvature* (MLSEC) operator.[9,10] Ridges and valleys of a gray-level image $L(x, y)$ are those image points that depict a lineal or locally anisotropic structure. Image ridge-like or valley-like structures (and creases in general) can be described as points of local extreme curvature of the image isointensity lines. Taking into account the relation between the curvature of isointenity curves and the parallelism of intensities gradients or "slopelines" (ones orthogonal to the others) the creasesness can be measured in an equivalent way as the divergence $\kappa = -\text{div}(\bar{\mathbf{w}})$ of the normalized gradient vector field, since divergence measures the degree of parallelism of the vector field, where:

$$\mathrm{div}(\mathbf{w}) = \frac{\delta w_x}{\delta x} + \frac{\delta w_y}{\delta y} \tag{8}$$

and, the normalized gradient vector field $\bar{\mathbf{w}}$ is defined as:

$$\bar{\mathbf{w}} = \frac{\mathbf{w}}{||\mathbf{w}||} \text{ if } ||\mathbf{w}|| > 0 \tag{9}$$

and 0 otherwise.

The pixels for which this operator response is below a fixed threshold of -1 are considered as valleys (black pixels at second row of Fig. 2), defining a binary image $I_A(x,y)$. The whole set $A = \{(x,y) : I_A(x,y) = 1\}$ of image positions where a valley has been detected is then considered as the face descriptor. Note that any direct Hausdorff measure, Eq. (4), between two sets A and B, that uses the average as *integral* function can be written as:

$$h(A,B) = \frac{1}{|A|} \sum_{a \in A} f^m \left(\min_{b \in B} d(a,b) \right) \tag{10}$$

$$= \frac{1}{\sum_x \sum_y I_A(x,y)} \sum_x \sum_y I_A(x,y) f^m \left(\min_{\{(i,j):I_B(i,j)=1\}} d((x,y),(i,j)) \right) \tag{11}$$

The rightmost term of Eq. (11) is equivalent to the distance transform, $DT(x,y;I_B)$, of the binary image $I_B(x,y)$. The distance transform $DT(x,y;I_B)$ of a binary image $I_B(x,y)$ is an image of the same size as $I_B(x,y)$ and defined so that the value of $DT(x,y;I_B)$ at each pixel position (x,y) is the Euclidean distance from the position (x,y) to the position (i,j) of the nearest nonzero pixel of the binary image $I_B(x,y)$. This image transformation can be efficiently computed in two passes over the image.[2] Both the binary images used to describe the face shape information and their distance transforms are shown on the second and third rows in Fig. 2. We will consider the binary images, their DT, and the result of the *best-match* function as the vectors $\mathbf{I_A}$, $\mathbf{D_A}$, and $\mathbf{f^m}(\mathbf{D_A})$. Then, Eqs. (10) and (11)

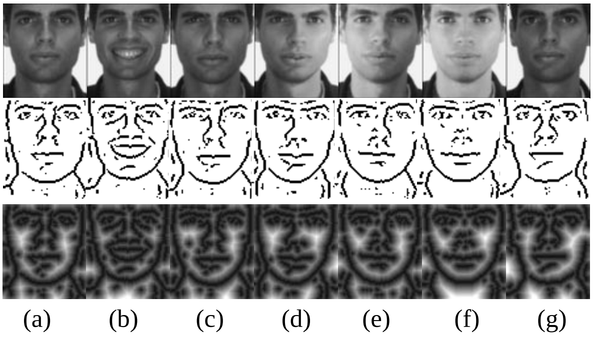


Fig. 2. Original cropped images (top row), their valley responses (second row) and distance maps (bottom row) of the valley responses.

can be rewritten as the inner product:

$$h(A,B) = \left\langle \frac{\mathbf{I_A}}{|\mathbf{I_A}|}, \mathbf{f^m}(\mathbf{D_B})) \right\rangle \tag{12}$$

4. Supervised Hausdorff Measures

In the case of AFR we have as many classes as different subjects to be recognized. And, as it has been explained before, each face image is represented by the set of image positions where features have been detected. Our purpose then, is to define a Hausdorff measure that allows us to classify images in an optimal way. In this case it is desired to obtain the *integral*, *symmetrizer*, and *best-match* function, so that the distance between two sets of different classes (subjects) become as large as possible and at the same time the distance between two sets of the same class (subject) become as small as possible. Although the three functions could be tuned to our classification problem, this paper presents only the results obtained when *best-match* function is modified, and average is used both as *integral* and *symmetrizer* function.

When two different frontal face images of the same subject are considered, it is worth to take into account that not all parts of the face suffer the same degree of transformation, in this sense, eyes and nose are more static, than mouth, chin or even eyebrows (mainly due to change of expression). The proposed *best-match* function has been designed so that the distance (deformation) contributions of each face point are normalized. In this way, the distortions measured in those places where a large degree of deformation (for different images of the same subject) is expected, are taken less into account than those features that remain almost constant for different images of the same subject. Given a set of samples, we use as a measure of elasticity of a face image point the within-class standard deviation of the distance transforms at that point. Formally, given a set of classes (subjects) $\mathbf{C} = \{\mathbf{C_1}, \mathbf{C_2}, \ldots, \mathbf{C_n}\}$, and $\mathbf{C_i} = \{\mathbf{D^i_1}, \mathbf{D^i_2}, \ldots, \mathbf{D^i_m}\}$ the DT of the different binary images of the class (subject) $\mathbf{C_i}$. Then, the result of applying the *best-match* function to a DT vector $\mathbf{D^i_B} \in \mathbf{C_i}$ is a vector $\mathbf{f^m}(\mathbf{D^i_B})$ of the same size than $\mathbf{D^i_B}$, where its nth component, $f^m_n(D^i_{B,n})$, is defined as:

$$f^m_n(D^i_{B,n}) = \frac{D^i_{B,n}}{\sigma_n} \tag{13}$$

where σ_n and $D^i_{B,n}$ are the nth components of the within class standard deviation vector and the nth component of the DT vector $\mathbf{D^i_B}$ respectively.

$$\begin{gathered} \sigma_n = \sqrt{\frac{1}{|C|} \sum_{\mathbf{C_i} \in \mathbf{C}} \frac{1}{|\mathbf{C_i}|} \sum_{\mathbf{D^i_j} \in \mathbf{C_i}} (D^i_{j,n} - \bar{D}^i_n)^2}, \text{ and,} \\ \bar{D}^i_n = \frac{1}{|\mathbf{C_i}|} \sum_{\mathbf{D^i_j} \in \mathbf{C_i}} D^i_{j,n} \end{gathered} \tag{14}$$

where $|.|$ denotes the cardinal of the set.

Fig. 3. Deformations (σ) measured for each source of variability. White values depict image points where distance measures for different images of the same subject have a high degree of variation. Black points depict pixels that remain static for different images of the same subject.

Figure 3 depicts the values of the normalization vector (σ) computed when all the sources of variation of the database are considered (rightmost image in Fig. 3). Each of the leftmost images at this figure, show the normalization vectors computed when only one isolated source of variation is considered. These correspond respectively from left to right and from up to down, to the normalization vectors computed when images change from neutral to smiling expression (conditions (a) and (b)), from neutral to angry expression ((a) and (c)), from diffuse to left illumination ((a) and (d)), from diffuse to right illumination ((a) and (e)), from diffuse to frontal illumination ((a) and (f)), and finally when differences are measured between images acquired in two different sessions ((a) and (g)).

5. Experimental Results

A set of 742 images from the AR-Face database,[11] have been used to test the system. These images correspond to 106 subject (seven shoots per subject labeled from (a) to (g) in Figs. 1 and 2, each label denoting a different acquisition condition). Eye location has been used in order to normalize the images in size, position, and orientation. Face images are then cropped and scaled to 78×68 pixels. Hit ratios are computed separately for each of the acquisition conditions so that the sensitivity of the recognition methods to different sources of variability can be observed. The sources of variation, respect to shoot (a), are: (i) changes of expression (smile (b), angry (c)), (ii)different illumination conditions (left light (d), right light (e), both left and right (f)), (iii) images taken in a second session two weeks later (g).

For each experimental condition, a pool of 212 images is constructed using the images with neutral expression, (a), and those with the considered source of variability (b) to (g). Each of these 212 images, is then considered as a test, using the remaining 211 images as gallery. For each test image, a hit is considered if the nearest gallery image and the test image belongs to the same subject. The hit ratio is then obtained dividing the number of hits per number of test images.

Table 1. Hit ratios obtained for different sources of variability (columns), and similarity measures (rows).

Method	(b)	(c)	(d)	(e)	(f)	(g)	Average
HD_V	17.92%	25.00%	19.34%	14.15%	6.13%	11.79%	15.72%
HF_V	14.62%	12.26%	11.79%	12.73%	15.09%	15.09%	13.60%
MHD_E	46.23%	61.32%	38.68%	54.24%	27.36%	8.96%	39.85%
MHD_V	48.58%	73.58%	81.60%	77.36%	49.06%	73.58%	67.29%
AHD_E	74.05%	67.92%	60.37%	54.24%	25.27%	68.39%	58.41%
AHD_V	78.77%	76.41%	88.68%	85.85%	69.34%	73.11%	78.69%

Table 2. Comparison between Average Hausdorff Distance (AHD) and the proposed Supervised Hausdorff Measures (SHM).

Method	(b)	(c)	(d)	(e)	(f)	(g)	Average
AHD_V	81.56%	80.14%	91.79%	86.93%	70.09%	73.96%	80.75%
$SHM_V(1)$	83.82%	75.61%	93.21%	89.01%	74.86%	80.38%	82.81%
$SHM_V(2)$	89.48%	80.14%	95.28%	90.71%	79.43%	84.43%	86.58%

Table 1 summarizes the results obtained with the different discussed methods (rows) when different sources of variability are considered (columns). The rightmost column shows the hit ratios obtained when the ratios for the different sources of variability are averaged. The results shown in this table have been obtained using valleys as features (except those labeled as MHD_E and AHD_E, that have been obtained using as shape feature descriptor the responses of a Canny edge detector), and Hausdorff Distance (HD_V), Average Hausdorff Distance (AHD_V), Modified Hausdorff Distance (MHD_V), Hausdorff Fraction (HF_V) (in this case recognition ratios have been computed for each tolerance radius d ranging from 1 to 15 pixels, the highest obtained recognition ratio is shown).

From the results shown in this table, it has to be noticed that valleys behave in a more robust way to illumination changes than edges (see columns (d) to (f)), and that the best recognition results are obtained when AHD is used as similarity criteria.

Further experiments have been conducted to validate the proposed Supervised Hausdorff Measure. In this case the face database has been split in training and test set. Half of the subjects have been randomly selected to form the training set, and their facial images have been used to tune the normalization parameters σ of the *best-match* function. The images of the 53 remaining subjects have been considered as test set. Average recognition hit ratio has been evaluated in the test set. This procedure has been cross-validated by repeating the experiment 20 times for each of the experimental conditions, and then averaging the obtained results. For comparison with the results reported in Table 1 average recognition hit ratios evaluated in the test sets have also been computed for AHD.

Table 2 summarizes the results obtained with the AHD and the proposed Supervised Hausdorff Measures (SHM_V). Results where supervised measures have been used are marked either as (1) or (2). In those indicated as (1), a unique elasticity normalization vector, σ, has been computed taking into account all the training set images acquired under different conditions (rightmost image at Fig. 3). Meanwhile for those marked as (2), a different normalization vector has been computed for each experimental condition using the training images of acquisition condition (a) and those of one of the remaining acquisition conditions ((b) to (g)) (leftmost images in Fig. 3).

The experimental results reported in this table show that the proposed supervised measures $SHD_V(1)$ and $SHD_V(2)$, outperform the recognition ratios obtained with the previously defined measures under all the acquisition conditions tested in our experiments.

6. Conclusions and Further Work

This paper is concerned with face recognition using only shape information. Hausdorff measures arise as a natural similarity measure for face shape information. A common framework has been introduced conveying all the Hausdorff based measures. The results presented in this paper shown that, on the tested database, image valleys are more robust (in terms of recognition ratios) to illumination changes than edges. Having a sample set of images under different acquisition conditions, allows us to compute a supervised Hausdorff-like measure. The recognition ratios obtained using this new measure overcome those obtained with previous Hausdorff measures reported in the literature. This work leaves a number of open possibilities that may be worth further research, among others it may be interesting to consider: (i) Design of optimal (in terms of classification) non-Euclidean between points distances d, *integral*, and *symmetrizer* functions. (ii) It may be worth to consider ridges in addition to valleys when computing similarity between images. (iii) Valleys responses to different scales could be taken into account.

Acknowledgments

This work has been partially funded by FUNDACION PROVIGO and grants TAP98-0618, CICYT TIC2000-0382 and FEDER 2FD97-1800.

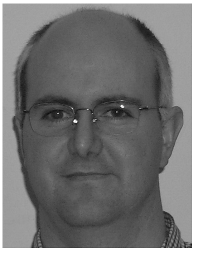
References

1. P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class-specific linear projection," *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 7 (1997) 711–720.
2. G. Borgefors, "Distance transformations in digital images," *CVGIP* **34**, 3 (1986) 344–371.
3. R. Chellappa, C. Wilson and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. IEEE* **83**, 5 (1995) 705–740.

4. M. Dubuisson and A. Jain, "A modified Hausdorff distance for object matching," *ICPR94*, 1994, pp. A:566–568.
5. R. Frischholz and U. Dieckmann, "Bioid: a multimodal biometric identification system," *Computer* **21**, 2 (2000) 64–68.
6. M. Grudin, "On internal representations in face recognition systems," *Patt. Recogn.* **33**, 7 (2000) 1161–1177.
7. D. Huttenlocher, G. Klanderman and W. Rucklidge, "Comparing images using the Hausdorff distance," *PAMI* **15**, 9 (1993) 850–863.
8. D. Huttenlocher, R. Lilien and C. Olson, "View-based recognition using an eigenspace approximation to the Hausdorff measure," *PAMI* **21**, 9 (1999) 951–955.
9. A. Lopez, D. Lloret, J. Serrat and J. Villanueva, "Multilocal creaseness based on the level set extrinsic curvature," *Comput. Vis. Imag. Underst.* **77** (2000) 111–144.
10. A. Lopez, F. Lumbreras, J. Serrat and J. Villanueva, "Evaluation of methods for ridge and valley detection," *PAMI* **21**, 4 (1999) 327–335.
11. A. Martinez and R. Benavente, "The ar face database," Technical Report 24, CVC, 1998.
12. B. Moghaddam, "Principal manifolds and Bayesian subspaces for visual recognition," *ICCV99*, 1999, pp. 1131–1136.
13. D. E. Pearson, E. Hanna and M. K., "Computer-generated cartoons," *Images and Understanding*, Cambridge University Press, 1990, pp. 46–60.
14. P. Phillips, M. Bone and D. Blackburn, "Facial report vendor test 2000: evaluation report," http://www.dodcounterdrug.com/facialrecognition/FRVT2000/documents.htm (2001).
15. P. Phillips, H. Moon, S. Rizvi and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *PAMI* **22**, 10 (2000) 1090–1104.
16. A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: a survey," *Patt. Recogn.* **25**, 1 (1992) 65–77.
17. B. Takacs, "Comparing face images using the modified Hausdorff distance," *PR* **31**, 12 (1998) 1873–1881.
18. D. Valentin, H. Abdi, A. O'Toole and G. Cottrell, "Connectionist models of face processing: a survey," *Patt. Recogn.* **27**, 9 (1994) 1209–1230.

Albert Pujol received the M.Sc. and Ph.D. in computer science at the University Autonoma of Barcelona (Spain) in 1996 and 2001, respectively, where he graduated cum laude. He has been teaching at the Computer Science Department of the Autonomous University of Barcelona from 1996 to 2001. He is a researcher at the Computer Vision Center.

His research interests include computer face recognition and biometrics.



Jose L. Alba received the M.Sc. and Ph.D. in telecommunications engineering from the University of Santiago and University of Vigo (Spain) in 1990 and 1997, respectively, where he graduated cum laude. He is an Associate Professor of discrete systems and speech and image processing at the University of Vigo, Spain.

His research interests include neural networks for classification applications, image segmentation, statistical pattern recognition, automatic speech and speaker recognition and biometrics.

Juan J. Villanueva received a B.Sc. in physics from the University of Barcelona in 1973 and a Ph.D. in computer science from the Universitat Autónoma de Barcelona in 1981. Since 1975, he has been teaching at the Computer Science Department, where he was appointed Professor in 1990. In 1986, Dr. Villanueva promoted the Computer Vision Center, and has been its Director since its foundation. He was co-founder and vice-president of AERFAI, which is the Spanish Chapter of IAPR; he is currently a member of its steering committee. Dr. Villanueva is a member of the IEEE, SPIE, and the IEEE Computer Society. He has participated in several Spanish and international boards of conferences and was co-Chairman of the International Conference on Pattern Recognition held in Barcelona in 2000.

His research interests comprise all aspects of computer vision, in particular recognition based on geometrical models and medical computer vision applications.