Recognition of coloured and textured images through a multi-scale neural architecture with orientational filtering and chromatic diffusion

M. Antón-Rodríguez *, F.J. Díaz-Pernas, J.F. Díez-Higuera, M. Martínez-Zarzuela, D. González-Ortega, D. Boto-Giralda

Department of Signal Theory, Communications and Telematics Engineering, Telecommunications Engineering School, University of Valladolid, Valladolid, Spain

A B S T R A C T

The aim of this paper is to outline a multiple scale neural model to recognise colour images of textured scenes. This model combines colour and textural information in order to recognise colour texture images through the operation of two main components: a segmentation component composed of the colour opponent system (COS) and the chromatic segmentation system (CSS); and a recognition component formed by an ARTMAP-based neural network with scale and orientation-invariance properties. Segmentation is achieved by perceptual contour extraction and diffusion processes on the colour opponent channels based on the human psychophysical theory of colour perception. This colour regions enhancement along with their local textural features constitutes the recognition pattern to be sent to the supervised neural classifier. The CSS accomplishes the colour region enhancement through a multiple scale loop of oriented filters and competition–cooperation mechanisms. Afterwards, the neural architecture performs an attentive recognition of the scene using those oriented filters responses and the chromatic diffusions. Some comparative tests with other models are included in order to prove the recognition capabilities of this neural architecture and how the use of colour information encourages the texture classification and the accuracy of the boundary detection.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

In biological vision, we can distinguish two main operating modes: pre-attentive and attentive vision. The first one performs a parallel and instantaneous processing which is independent of the number of patterns being processed, thus covering a large region of the visual field. Attentive vision, nevertheless, acts over limited regions of the visual field (small aperture) establishing a serialised search by means of focal attention [19]. The proposed model works on the pre-attentive and attentive modes: pre-attentive segmentation and attentive recognition. In the pre-attentive process, the network consistently processes colour and textural information for enhancing regions and extracting perceptual boundaries to form the segmented image. In the attentive mode, the model merges the textural information and the intensity of the region enhancement in order to punctually recognise scenes including complex textures, both natural and artificial.

The skill of identifying, grouping and distinguishing among textures and colours is inherent to the human visual system. In the last few years many techniques and models have been proposed in the area of textures and colour analysis [6], resulting in a detailed characterisation of both parameters as well as certain rules that model their nature. Measurement of colour and texture in combination, rather than colour or texture alone, would provide better discriminating power. This topic is attracting many researchers nowadays. There are many proposals for systems that combine both colour and texture not only at segmentation but also at recognition level. Mirmehdi and Petrou [21] described an approach to the segmentation of colour image textures. This system proposed a multiple scale representation of the texture based on human psychophysical measurements of colour appearance. In this work, they showed that colour attributes which characterise a colour texture at a certain resolution may be entirely different from the colour attributes that characterise the same texture at another resolution. Thai and Healey [27] proposed measuring colour texture by embedding the Gabor filters into an opponent colour representation. With an analogous theorising, Hoang et al. [13] used Gabor–Gaussian filters to generate colour texture measurement for the segmentation. Both methods provide a useful structural representation for colour texture through a bank of filters with different filtering parameters. The processing of multiple filters with different scales is most suited for the segmentation of colour textures. In this sense, the neural architecture presented hereby in this paper implements a multi-scale processing using Gabor filtering, and it characterises colour through neighbouring transformations of opponent chromatic

---

signals of short, middle and long wavelength, a modelling more kindred to the human visual system. The neural architecture presented hereby in this paper, implements a multi-scale processing and characterises colour through neighbouring transformations of opponent chromatic signals of short, middle and long wavelength, a modelling more kindred to the human visual system.

Colour and texture have also been used by researchers in the recognition of textured scenes. Paschos [24] revealed that incorporating colour into a texture analysis and recognition scheme can be very beneficial. This paper compares RGB, with Lab and HSV in terms of their effectiveness in colour texture analysis. Zhong and Jain [29] used local texture and colour features to localise objects. The texture and colour features are directly extracted from the discrete cosine transform (DCT) compressed domain thereafter using a deformable template matching method to match the query shape to the image edges at the locations which possess the desired texture and colour attributes. Permuter et al. [25] defined Gaussian mixture models (GMMs) of coloured texture on several feature spaces and analysed the performance of these models in various classification tasks. They compared supervised classification results for different choices of colour and texture features and explored the best set of features and the best GMM configuration for this task. A maximum likelihood decision is used for this classification. Jain and Healey [18] introduced a representation for colour texture using monochrome and opponent features computed from Gabor filter outputs. The monochrome features are computed from the spectral bands independently while the opponent features combine information across different spectral bands at different scales, using three scales and four orientations to determine the distance from another texture. A texture test was classified as an instance of the database texture with which it has the smallest distance. Martin et al. [22] detected and localised boundaries using characteristic changes in brightness, colour, and texture. To combine the information from these features they trained a classifier using human labelled images as ground truth. All these works show the importance of colour in texture identification. This paper also includes an analysis in this sense. Tests with and without using colour information are performed, achieving best results in the first case.

From a biological point of view, the filtering using different spatial frequencies and orientations and the parallel processing in opponent colour channels are active parts of the mammalian visual system when perceiving textured scenes [30]. Inspired in such facts, the neural model architecture described in this work, which we called ChromARTex, extracts both colour and textural features from a scene through a multiple scale processing of the colour-opponent channels, segments it into textural regions and brings this information to an ART-based (adaptive resonance theory) classifier, which categorises the textures using a biologically motivated learning algorithm.

The proposed architecture is based on later versions of BCS/FCS (boundary contour system/feature contour system) neural model [10,23], and on ARTMAP recognition architectures [1,2]. The BCS/FCS model, within the theory of neural dynamics of the visual perception, suggests a neural dynamics for perceptual segmentation of monochromatic visual stimuli and offers a unified analysis process for different data referring to monocular perception, grouping, textural segmentation and illusory figures perception. The BCS system obtains a map of image contours based on contrast detection processes, whereas the FCS performs diffusion processes with luminance filling-in within those regions limited by high contour activities. Consequently, regions that show certain homogeneity and are globally independent are intensified. In recent publications, Grossberg et al. focused on the modelling

of retinal processes in order to solve the problem of consistent interpretation of surface lightness under varying illumination. This biological process known as "anchoring" has been thoroughly studied and modelled by those researchers. Hong and Grossberg [14] developed a lightness perception model that simulates such a biological visual process as well as other processes like discounting the illuminant and lightness constancy. Furthermore, this model is able to process natural colour images under variable lighting conditions and is compared with the Retinex model. The model proposed in [9] clarifies how retinal processing stages achieve light and spatial contrast adaptation. This research group has, in addition, developed a shunting system called DISCOV (Dimensionless Shunting Colour Vision), which models a cascade of primate colour vision cells [4]. As it can be observed, this group has very interesting approaches over retinal stages.

In this manner, the main improvement introduced to the last BCS/FCS model hereby in this paper, resides in offering a complete colour image processing neural architecture from the input RGB image up to the segmentation level. In order to do this, the neural architecture performs processing stages for extracting contours and enhancing the homogeneous areas in a colour image, following analogous behaviours to those of the early mammalian visual system. This adaptation has been performed by trying to preserve the last BCS/FCS model perceptual qualities. ChromARTex establishes a parallelism among different visual information channels and models the physiological behaviour of the visual system processes. Therefore, the envisaged region enhancement is based on the feature extraction and perceptual grouping of region points with similar and distinctive values of luminance, colour, texture and shading information. This paper shows how this colour information helps to improve identifying textured regions. Hence, it participates in the recognition process.

The adaptive categorisation and predictive theory is called adaptive resonance theory, ART. ART models are capable of stably self-organising their recognition codes using either unsupervised or supervised incremental learning [2]. ARTMAP theory extends the ART designs to include supervised learning. Fuzzy ARTMAP and default ARTMAP architectures [1] fall into this supervised theory. In ARTMAP architectures, the ART chosen categories learn to make predictions which take the form of mappings to the names of output classes. And thus many categories can map the same output name. ARTMAP architectures have been used in many different applications for classifying and recognition. Carpenter et al. [3] used a default ARTMAP architecture for producing a self-organising expert system that discovers hierarchical knowledge structures. This system was tested with Monterey and Boston images with good results. On the categorisation level, ChromARTex advances an ARTMAP-based neural classifier with two essential differences in comparison to ARTMAP, the inclusion of an orientational-invariance process and a faster categorisation and labelling method.

Grossberg and Williamson [12] proposed a neural system, ARTEX, for recognising textured scenes. ARTEX is composed of a BCS filter, a FCS stage and a Gaussian ARTMAP classifier. In this work, ARTEX and the model proposed by Greenspan [7] were compared in classifying Brodatz textures obtaining good results. We include in our work a comparison with ARTEX and Greenspan's architecture in order to appraise the performance of our model and the colour importance in texture recognition. As ARTEX and ChromARTex models share the theoretical base, the importance of adding colour processing to texture recognition can be more deeply evaluated.

The outline of this paper is as follows. In Section 2, each of the stages composing the architecture will be explained. Afterwards, Section 3 studies its performance over input images presenting complex textures in order to, in Section 4, establish the

conclusions of the analysis and finally assess the validity of the model depicted here.

2. Proposed neural model

The architecture of the ChromARTex model (Fig. 1) is composed of two main components: a colour segmentation module and a recognition module. The first component consists of two systems called colour opponent system (COS) and chromatic segmentation system (CSS). The recognition module (RM) is made up by a feature smooth stage, and an orientation-invariance neural network classifier. Chromatic extension and modifications incorporated to the BCS/FCS model are fully presented in the CSS module, since COS module has no relation to it.

The COS module transforms the chromatic components of the input signals (RGB) into a bio-inspired codification system, made up of various opponent chromatic channels and an achromatic channel. In order to do this, the COS firstly calculates the activations of the long (L), middle (M), and short (S) wavelength of the retinal cones, and then it generates the opponent processes of those signals, corresponding to the L–M and S–(L+M) opponencies.

Resulting signals from COS are used as inputs for the CSS module where the contour map extraction and two intensified region images, corresponding to the enhancement of L–M and S–(L+M) opponent chromatic channels, are generated in a multiple scale processing according to various perceptual mechanisms (perceptual grouping, illusory contours, discounting the illuminant and emergent features). The two enhanced images along with the textural response from the simple cells form the features pattern that will be sent to the recognition module where the fuzzy ARTMAP-based network generates a context-sensitive classification of loca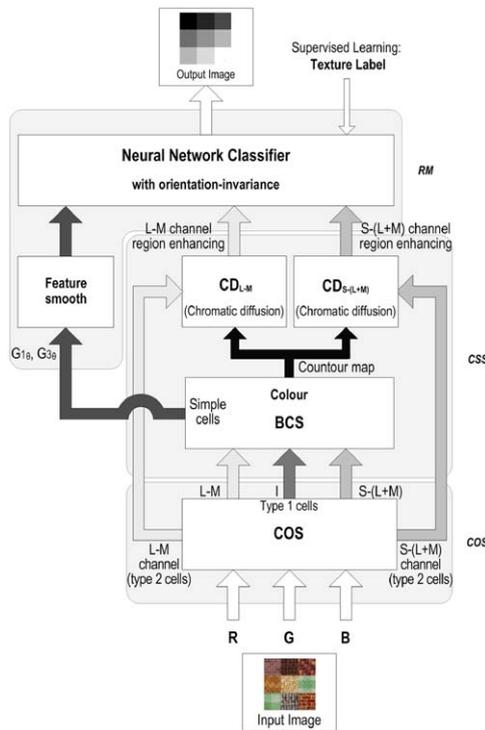l patterns. The final output of the proposed neural architecture is a texture class prediction image where each point is associated to the texture class label which it belongs to.

3. Colour opponent system (COS)

The COS module of ChromARTex performs colour opponent processes based on opponent mechanisms that are present on the retina and on the LGN (lateral geniculate nucleus) of the mammalian visual system [15]. Fig. 2 shows the detailed structure of the COS system.

The COS module includes opponent channels parallel processing. Firstly, it calculates the activations of the long (L), middle (M), and short (S) wavelength of the retinal cones (1) [16], and the luminance (I) (2) [6] and Y ($Y = L+M$) signals. Next, it generates the L–M and S–(L+M) opponent channels according to two types of cells we called "type 1 cells", which initiate the contour extraction, and "type 2 cells", which trigger the double opponent diffusion process.

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.293 & 0.603 & 0.104 \\ 0.134 & 0.704 & 0.162 \\ 0.046 & 0.099 & 0.854 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

$$I = 0.299R + 0.587G + 0.114B \tag{2}$$

Type 1 opponent cells are highly sensitive and precise to contrasts and light-darkness contours, so they initiate the process for image contour and texture detection. Type 2 opponent cells initiate the double opponent process that take place in the chromatic diffusion stage. The double opponent mechanisms are fundamental in human visual colour processing [15].

Type 1 cells are modelled through two centre-surround multiple scale competitive networks, so they form the ON and OFF channels composed of ON-centre OFF-surround and OFF-centre ON-surround competitive fields, respectively. The ON channel processing is in charge of enhancing the information of regions with higher intensity (or chromatic activity) in relation to their surroundings, while the OFF channel enhances regions with lower intensity (chromatic value) than their surroundings. These competitive processes establish a gain control network over the inputs from chromatic and luminance channels, maintaining the sensibility of cells to contrasts, compensating variable illumination, and normalising image intensity [11]. The following
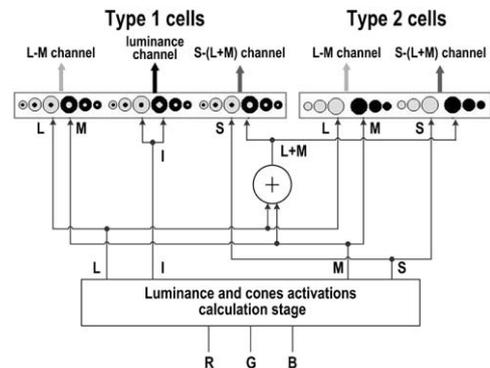

Fig. 1. Proposed model architecture. It is composed of: the colour opponent system (COS), the chromatic segmentation system (CSS) comprised of the colour boundary contour system (Colour BCS) and two chromatic diffusions (CD), and the recognition module (RM).

Fig. 2. Detailed COS module structure: on the left it shows type 1 cells whereas on the right, elements corresponding to type 2 opponent cells. Both types include ON and OFF cells for each of the opponencies and luminance channel (the latter only for type 1).

equations model the ON and OFF channels, respectively:

$$y_{ij}^{g+} = \left[\frac{AD^+ + BS_{ij}^c - CS_{ij}^{sg}}{A + S_{ij}^c + S_{ij}^{sg}} \right]^+ \tag{3}$$

$$y_{ij}^{g-} = \left[\frac{AD^- + BS_{ij}^{sg} - CS_{ij}^c}{A + S_{ij}^{sg} + S_{ij}^c} \right]^+ \tag{4}$$

where A, B, C and D are model parameters, $[w]^+ = \max(w,0)$ and

$$S_{ij}^c = \sum_{pq} e_{i+p,j+q}^c G_{pq}^c$$

$$S_{ij}^{sg} = \sum_{pq} e_{i+p,j+q}^{sg} G_{pq}^{sg} \tag{5}$$

with e^c as central signal, e^{sg} as peripheral signal ($e^c = L_{ij}$ and $e^{sg} = M_{ij}$ for $L-M$ opponency, $e^c = S_{ij}$ and $e^{sg} = Y_{ij}$ for $S-(L+M)$ opponency, and $e^c = e^{sg} = I_{ij}$ for the luminance channel) and the superscript $g = 0,1,2$ with suitable values for the small, medium and large scales; subscripts i, j, p, q are position indexes. The weight functions have been defined as normalised Gaussian functions for central (G^c) and peripheral (G^{sg}) connectivity:

$$G_{pq}^d = \frac{1}{\sigma_d \sqrt{2\pi}} \exp\left(-\frac{p^2 + q^2}{2\sigma_d^2} \right) \tag{6}$$

where σ_d is the standard deviation. Values used in the tests performed are $\sigma_c = 0.3$ for the central signal, and $\sigma_{ss} = 0.5$, $\sigma_{sm} = 1.0$, $\sigma_{sl} = 1.8$ for the peripheral signal in each of the scales.

The receptive fields of type 2 cells are composed of a unique central region in which two opponent colour processes occur, corresponding to $L-M$ and $S-(L+M)$ channels (see Fig. 2). Each opponent process is modelled by a multiplicative competitive central field, simultaneously presenting an excitation and an inhibition caused by different types of cone signals (L, M, S and Y). These processes are applied over the three spatial scales in the multiple scale model shown. Eqs. (3) and (4) also model the behaviour of these cells considering $B = C$, $\sigma_{c=\sigma_{sg}}$, and the input signals to the opponent process as $e^c = L_{ij}$ and $e^{sg} = M_{ij}$ for $L-M$ opponency and $e^c = S_{ij}$ and $e^{sg} = Y_{ij}$ for $S-(L+M)$ opponency. Values modelling these cells in the tests performed are $\sigma_s = 0.3$ and $\sigma_m = 0.8$, and $\sigma_l = 1.6$.

As it can be noted, not all possible opponency combinations have been modelled, avoiding, for example, cells like M^+-L^- or $(L+M)^+-S^-$ and their opposites; nevertheless, the information made available from the included combinations seems to be sufficient to continue with coherent image segmentation. Those combinations are also suitable inputs for the double opponent cells included in the next steps.

4. Chromatic segmentation system (CSS)

As previously mentioned, the chromatic segmentation system bases its structure on the modified BCS/FCS (boundary contour system/feature contour system) model [10,23], adapting its functionality to chromatic opponent signals, for colour image processing. Fig. 3 shows the detailed structure of the CSS module.

Next, the most significant changes performed upon the latter version of the BCS/FCS model in order to process colour information, are detailed:

- Contrast enhancement stage has been suppressed since it has been previously included in the COS type 1 opponent cells.
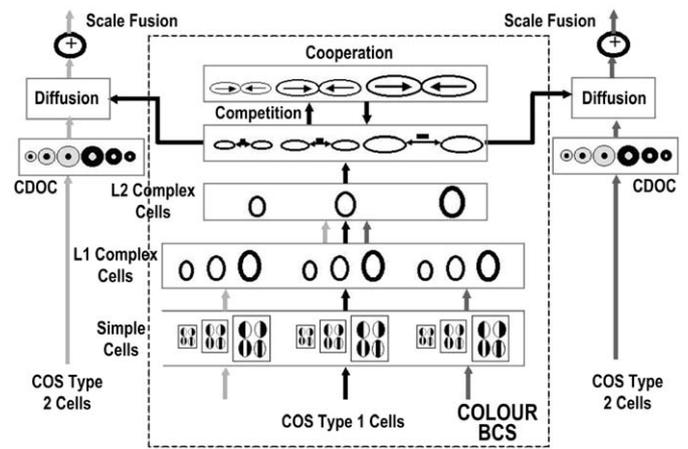- New type 2 cells are included to initiate double opponency for diffusion processes.

Fig. 3. Detailed structure of the chromatic segmentation system (CSS) based on the BCS/FCS model.

- Simple cell stage is applied on the proposed opponent channels (chromatic and luminance) in parallel. Hence, contours are extracted from each of the colour and luminance channels.
- The original simple cells' receptive fields have been substituted by Gabor filters.
- Complex cell stage has been divided into two layers: the first one fuses the information related to each Gabor filter and the second combines the outputs of three opponent channels to obtain a unified contour map.
- The FCS module adds an initial stage that simulates the behaviour of chromatic double opponent cells (CDOC stage). This stage uses the outputs of COS type 2 opponent cells as inputs and it constitutes a previous step for the diffusion stage.
- The FCS module has been duplicated in order to obtain two different diffusions, one for each chromatic opponency.
- Diffusion process is performed over inherent features of the colour regions, different wavelength opponencies, unlike BCS/FCS model which uses contrast signals. Inherent attributes can participate into the recognition process while contrast signal depends on the surrounding regions, which does not suit best for recognition.
- Diffusions are performed over double opponencies and not over simple like the original model.

The CSS module consists of the Colour BCS stage and two chromatic diffusion stages, processing one chromatic channel each. These stages are described below.

4.1. Colour BCS stage

The Colour BCS stage (see Fig. 3) constitutes our colour extension of the BCS model. It processes visual information from the three parallel channels, luminance and two chromatic channels, to obtain a unified contour map. The Colour BCS module has two differentiated phases: the first one (simple and complex cells) extracts real contours from the output signals of the COS module and the second is represented by a competition and cooperation loop, in which real contours are completed and refined, thus generating contour interpolation and illusory contours. Colour BCS presents characteristics such as perceptual grouping, emergent features and discounting the illuminant.

The achieved output coming from the competition stage is a contour map of the original image. This output is transmitted to the diffusion stages where it will act as a control signal serving as a barrier in chromatic diffusions. That is the reason why the

chromatic diffusion signals will facilitate detecting region transitions later in recognition stage.

4.1.1. Simple cells

Simple cells respond to variations on the textural and colour features of visual stimuli. The simple cells are sensitive to both orientation and scale characteristics. Hence, textural and colour feature variations for each orientation and scale resulting in this stage will constitute the main part of the pattern used to accomplish the model's objective, that is, the texture recognition.

This stage is modelled through two opposite pairs of Gabor filters (see Fig. 3), due to their high sensibility to orientation, spatial frequency and position [5]. Their presence on the simple cells situated at V1 area of visual cortex has been proved [26]. This receptive fields are detailed by

$$G_{ijk}^{(1)g} = \exp\left(\frac{-(i'/\lambda)^2 + j'^2)}{2\sigma_g^2}\right)\sin(2\pi F_g i')$$

$$G_{ijk}^{(3)g} = \exp\left(\frac{-(i'/\lambda)^2 + j'^2)}{2\sigma_g^2}\right)\cos(2\pi F_g i') \qquad (7)$$

where $(i',j') = (i\cos k + j\sin k, -i\sin k + j\cos k)$ represents a coordinate transformation with a rotation angle $k \in [0°, 180°]$. With $G^{(2)g} = -G^{(1)g}$ and $G^{(4)g} = -G^{(3)g}$ we obtained two pairs of Gabor filters with opposite polarity.

This function responds to a sinusoidal wave with F_g as spatial frequency, modulated by a Gaussian with aspect ratio λ and a standard deviation σ_g for each of the spatial scales (g). Odd cells ($G^{(1)}$ and $G^{(2)}$) are responsible for the detection of contours derived from contrasts with light–dark or dark–light transitions following the orientation k, while even cells ($G^{(3)}$ and $G^{(4)}$) are responsible for the detection of light level line or dark level line segments with orientation k surrounded by dark or light regions, respectively. The inclusion of $G^{(3)}$ and $G^{(4)}$ fields is justified due to the importance of these profiles in texture detection [20]. Values used in the tests performed for the three scales are $\sigma_s = 8.0$, $\sigma_m = 12.0$, and $\sigma_l = 15.0$.

The simple cell stage acts over chromatic and luminance channels separately, and also over each of the three envisaged spatial scales. The use of this multiple scale model reflects, at this stage, some of its main features, presenting reactive cells both to abrupt contrasts and low image gradients. These attributes are required to detect and recognise textures. The equations governing this stage are

$$r_{ijk}^{(d)g} = [(y_{ij}^{+g} - y_{ij}^{-g}) \otimes G_{ijk}^{(d)g}]^+ \qquad (8)$$

where $(y_{ij}^{+g} - y_{ij}^{-g})$ constitutes the subtraction of the ON and OFF type 1 cells, $d = 1,2,3,4$ symbolises the Gabor function used, g indicates the spatial scale ($g = 0,1,2$), subscripts i, j are position indexes and subscript k the orientation, $\otimes$ represents a two-dimensional convolution, and $[w]^+ = \max(w,0)$ is a half-wave rectifier.

4.1.2. Complex cells

The complex cell stage, using two cellular layers, fuses information from simple cells giving rise to a map which contains real contours for each of the three scales used (see Fig. 3).

The first layer of cells is in charge of combining responses from different Gabor filters at each opponent channel, L–M, S–Y, and luminance at their three scales. The second complex cell layer fuses information from the three opponent channels (L–M, S–Y, luminance) generating a final map of real contours.

4.1.3. Competition stage

Detected real contours are passed into a cooperative–competitive loop, as it is shown in Fig. 3. This nonlinear feedback network detects, regulates, and completes boundaries into globally consistent contrast positions and orientations, while it suppresses activations from redundant and less important contours, thus eliminating image noise.

The main implication of this cooperative–competitive interaction loop is the illusory contour extraction. The loop completes the real contours in a consistent way generating, as a result, the illusory contours. In order to achieve this feature it makes use of a short-range competition, and a long-range cooperation stage [10,23].

Boundary competition is applied at two image domains: spatial position and orientation. In this manner, the location and orientation of each of the contour's points are more precisely determined. Furthermore, as stated in [23], modelled hypercomplex cells are equipped with a property named "endstopping", which makes boundaries more vigorous at the end of a line than they are in the middle. This property allows the line ends to participate into illusory contour generation processes (quite a common feature in textural perception). The equations used in this stage are taken from the BCS model [10].

4.1.4. Cooperation stage

Cooperation is carried out by dipole cells, which have been placed just before cortical cells in V2 area. These cells have been used to model processes such as illusory contour generation, neon colour spreading or texture segregation [8].

Dipole cells act like long-range statistical AND gates, providing active responses if they perceive enough activity over both dipole receptive fields lobes (left and right) (see Fig. 4). Thus, this module performs a long-range orientation-dependent cooperation in such a way that dipole cells are excited by collinear (or close to collinearity) competition outputs and inhibited by perpendicularly oriented cells. This property is known as spatial impermeability and it prevents boundary completions towards regions containing substantial amounts of perpendicular or oblique contours. As in the competition stage, the equations for the cooperation stage have been taken from the original model [10].

Between competition and cooperation stages, a loop is established to strengthen and interpolate contours until a stable state is reached. Cooperation–competition loop is applied in parallel over three spatial scales.

4.2. Chromatic diffusion stage

As mentioned above, the chromatic diffusion stage has undergone changes that entailed the introduction of chromatic double opponency cells (CDOC), resulting in a new stage in the segmentation process.
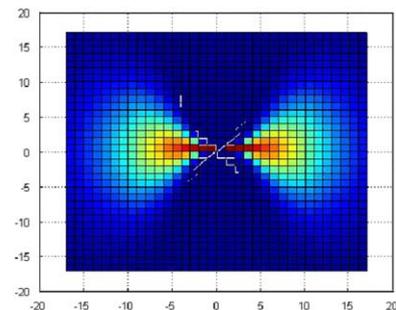

Fig. 4. Receptive field used to model dipole cells.

4.2.1. Chromatic double opponent cells (CDOC)

CDOC stage models chromatic double opponent cells (see Fig. 3). In human visual system, double opponency occurs in visual striate cortex cells, contained in blobs [15]. The model for these cells is the same receptive field as COS type 1 opponent cells (centre-surround competition), but their behaviour is quite a lot more complex since they are highly sensitive to chromatic contrasts. Double opponent cell receptive fields are excited on their central region by COS type 2 opponent cells, and they are inhibited by the same cell type. Double opponency is applied both in L–M and S–Y channels. This is to say, we apply a greater sensibility to contrast as well as a more correct attenuation toward illumination effects, therefore bringing a positive solution to the noise-saturation dilemma.

The mathematical pattern that governs the behaviour of chromatic double opponent cells is the one defined by Eq. (3) and successive equations, by varying only their inputs. These inputs are now constituted by the outputs of the COS type 2 opponent cells for each chromatic channel, that is, $e^c = e^{sg} = (L^+ - M^-)_{ij}$ for $L^+ - M^-$ channel, $e^c = e^{sg} = (L^- - M^+)_{ij}$ for $L^- - M^+$, $e^c = e^{sg} = (S^+ - Y^-)_{ij}$ for $S^+ - Y^-$, and $e^c = e^{sg} = (S^- - Y^+)_{ij}$ for $S^- - Y^+$. Values used in the tests performed are $\sigma_c = 0.3$ for the central signal, and $\sigma_{ss} = 1.2$, $\sigma_{sm} = 5.2$, $\sigma_{sl} = 10.8$ for the peripheral signal in each of the scales.

4.2.2. Diffusion

Diffusion stage performs four nonlinear and independent diffusions for L–M (ON and OFF) and S–Y (ON and OFF) chromatic channels. These diffusions are controlled by means of a final contour map obtained from the competition–cooperation loop (in particular, from the competition hyper-complex cells) while the outputs of CDOC are the signals being diffused (see Fig. 3).

At this stage, each spatial position diffuses its chromatic features in all directions except those in which a boundary is detected. When boundary signals take part, they inhibit diffusion obtaining differentiated activities at each of their sides (thus separating regions with different features). Contours, therefore, behave as a form-sensitive mesh that is called "boundary web" [23], being reactive to boundaries, textures, shading, and colours, while smoothing statistically irrelevant noise. By means of this process, image regions that are surrounded by closed boundaries tend to obtain uniform chromatic features, even in noise presence, and therefore producing the enhancement of the regions detected in the image, which favours the recognition. The equations that model the diffusive filling-in can be found in [10].

Again, diffusion is independently performed over the three spatial scales in an iterative manner, obtaining new results from previous excitations, simulating a liquid expansion over a surface. In our tests, we have used 800 iterations to reach a stabilised state.

4.2.3. Scale fusion

Scale fusion constitutes the last stage of pre-processing before recognition (see Fig. 3). On each scale, signals from opposite diffusion channels are first combined and then these results are fused again with the information from other spatial scales. The weighted coefficients depend on the noise level of the input image as well as on the envisaged use of the segmentation (in relation to the perception intention: a lot of detail or more general). With noiseless inputs, smaller scales obtain detailed results while greater noise levels require larger scales.

This output stage forms part of the recognition stage input pattern. These two components (L–M and S–$(L+M)$) are essential as they bring the model in all the boundary-surface information.

5. Recognition module

The recognition process takes place in two steps: a feature-smoothing stage and an orientational invariant neural network classifying stage (see Fig. 1). The input to the neural network classifier is a pattern formed by the aforementioned smoothed output and the colour information extracted from the diffusion stage (chromatic channels).

The classifier is trained using supervised learning and shifting input patterns to achieve orientational invariant texture recognition. A new supervised texture labelling process maps a colour and textural pattern onto a texture category. Final recognition process output provides a texture category prediction for every point on the image.

5.1. Texture feature smooth stage

In order to shape the patterns, the full-wave rectifier responses coming from two simple cell filters are used: the anti-symmetric light-dark receptive field ($G^{(1)}$) and the symmetric receptive field with central excitation ($G^{(3)}$) (see (7)). With them, we used the three spatial scales ($g = 0,1,2$ for small, medium and large scales) and N orientations ($n = 0,1,\ldots,N-1$ for angles $k \in [0°, 180°]$), thus, obtaining a $6 \cdots N$-dimensional textural vector. Due to the high spatial variability shown in Gabor's filters response, a smooth stage for each scale is proposed through a Gaussian kernel (6) convolution with a σ^g_{smooth} deviation for each scale in all orientations used. Values modelling this smoothing in the tests performed are $\sigma^0_{smooth} = 1.5$, $\sigma^1_{smooth} = 2.5$, and $\sigma^2_{smooth} = 6$.

The following equation models this stage:

$$t_{ijk}^{(1,3)g} = |r_{ijk}^{(1,3)g}| \otimes G^g \tag{9}$$

where $|r_{ijk}|$ is the simple cell output in position (i,j) and orientation k, G is the Gaussian kernel, $g = 0,1,2$ depending on the scale. So, input pattern to the neural stage is

$$\mathbf{P}_{ij} = [t_{ij0}^{(1)0}, t_{ij0}^{(1)1}, t_{ij0}^{(1)2}, t_{ij0}^{(3)0}, t_{ij0}^{(3)1}, t_{ij0}^{(3)2}, \ldots,$$
$$t_{ijN-1}^{(1)0}, t_{ijN-1}^{(1)1}, t_{ijN-1}^{(1)2}, t_{ijN-1}^{(3)0}, t_{ijN-1}^{(3)1}, t_{ijN-1}^{(3)2}, V_{ij}^{L-M}, V_{ij}^{S+(L-M)}]$$

where $V_{ij}^{L-M}, V_{ij}^{S+(L-M)}$ are the diffusion stage outputs (L–M and S–$(L+M)$ channels).

5.2. Neural network classifier

The neural network used in this stage is based on ARTMAP theory [1,2], though several variations are included (see Fig. 5):

- F_2 activation is done through the maximum activity criterion but in a multi-pattern setting.
- The vigilance criterion is evaluated over the maximum activity generator.
- Only those F_2 nodes linked to the supervised label are activated.

In the pattern categorisation process some orientational invariance are generated by means of the group displacement of components following the pattern's existing orientations. Using N different orientations, these displacements will produce N patterns, $\mathbf{P}_{ijn}$ ($n = 0, 1,\ldots, N-1$) (see Eq. (10)), taking part in the activation of the categories level and the reset control of the neural classifier. The two last components coming from diffusion stage do not participate in this shifting. Thanks to these invariances, it is achieved that the same texture pattern may be recognised from different angles.
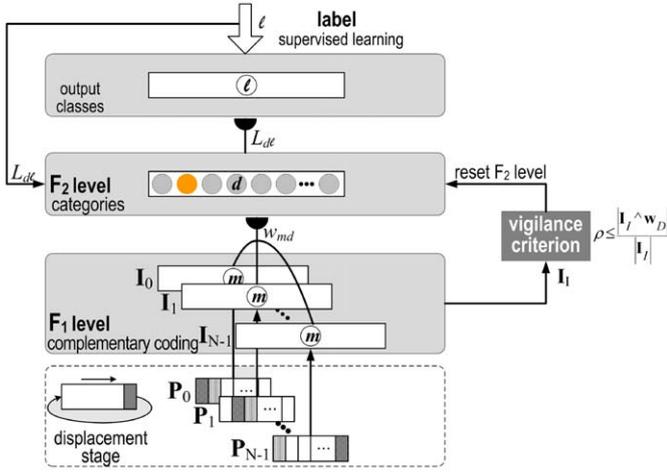
Fig. 5. Detailed neural classifier structure based on ARTMAP networks.

$$\mathbf{P}_{ij0} = [t_{ij0}^{(1)0}, t_{ij0}^{(1)1}, t_{ij0}^{(1)2}, t_{ij0}^{(3)0}, t_{ij0}^{(3)1}, t_{ij0}^{(3)2}, \dots,$$

$$t_{ijN-1}^{(1)0}, t_{ijN-1}^{(1)1}, t_{ijN-1}^{(1)2}, t_{ijN-1}^{(3)0}, t_{ijN-1}^{(3)1}, t_{ijN-1}^{(3)2}, V_{ij}^{L-M}, V_{ij}^{S+(L-M)}]$$

$$\mathbf{P}_{ij1} = [t_{ijN-1}^{(1)0}, t_{ijN-1}^{(1)1}, t_{ijN-1}^{(1)2}, t_{ijN-1}^{(3)0}, t_{ijN-1}^{(3)1}, t_{ijN-1}^{(3)2}, \dots,$$

$$\dots t_{ijN-2}^{(1)0}, t_{ijN-2}^{(1)1}, t_{ijN-2}^{(1)2}, t_{ijN-2}^{(3)0}, t_{ijN-2}^{(3)1}, t_{ijN-2}^{(3)2}, V_{ij}^{L-M}, V_{ij}^{S+(L-M)}]$$

$$\mathbf{P}_{ijN-1} = [t_{ij1}^{(1)0}, t_{ij1}^{(1)1}, t_{ij1}^{(1)2}, t_{ij1}^{(3)0}, t_{ij1}^{(3)1}, t_{ij1}^{(3)2}, \dots,$$

$$t_{ij0}^{(1)0}, t_{ij0}^{(1)1}, t_{ij0}^{(1)2}, t_{ij0}^{(3)0}, t_{ij0}^{(3)1}, t_{ij0}^{(3)2}, V_{ij}^{L-M}, V_{ij}^{S+(L-M)}] \tag{10}$$

The network (see Fig. 5) has two levels of neural layers, the input level F_1 with N layers and complementary coding, and the categorisation level F_2, where its output is linked to the texture label $(l = 0, 1, \dots, N_l - 1)$ by a linking weight L_{dl} defined by Eq. (11). A node of F_2 represents one category formed by the network and it is characterised by its weight vector $\mathbf{w}_{md}$ with $\{m = 0, 1, \dots, M - 1\}$ and $\{d = 0, 1, \dots, N_c - 1\}$ where M is the layer dimension of the F_1 level, $M = (6*N+2)*2$ (complementary coding), and N_c the number of committed nodes in F_2.

$$L_{dl} = \begin{cases} 1 & \text{if category } d \text{ is linked to texture } l \\ 0 & \text{in other case} \end{cases} \tag{11}$$

5.2.1. Learning process

The N vectors generated by the displacement process are changed into their complement coding form:

$$\mathbf{I}_n = (\mathbf{P}_{ijn}, 1 - \mathbf{P}_{ijn}) \tag{12}$$

with $n = 0, 1, \dots, N - 1$ and $\mathbf{P}_{ij}$ as it is pointed out in (10). The activity patterns on layer F_1 is set equal to $\mathbf{I}_n$. For a texture class with a supervised label l $(l = 0, 1, \dots, N_l - 1)$, the choice function is then calculated by Eqs. (13) and (14) for each committed node of F_2, for those committed nodes linked to the same texture label as the ones being processed, that is, if $L_{dl} = 1$. For F_2 nodes linked to other texture class (i.e. $L_{dl} = 0$) the activity of the F_2 committed node is reset because a supervised learning is done. Initially, all the F_2 nodes have a null linking weight, $L_{dl} = 0$.

This choice function is defined by

$$T_{nd} = \frac{|\mathbf{I}_n \,{}^\wedge \mathbf{w}_d|}{\alpha + |\mathbf{w}_d|} \tag{13}$$

$$T_d = \begin{cases} \max\{T_{nd} : n = 0, \dots, N - 1\} & \text{if } L_{dl} = 1 \\ 0 & \text{if } L_{dl} = 0 \end{cases} \tag{14}$$

where $|\cdot|$ is the L_1 norm of the vector, $^\wedge$ is the fuzzy AND operator $((\mathbf{p}{}^\wedge\mathbf{q})_i \equiv \min(p_i, q_i))$ and $\alpha > 0$ is the choice parameter, usually chosen close to zero for a good performance. In all our test simulations, we have used $\alpha = 0.05$.

A winner takes all decisions used in F_2 and node D is the one selected (see (15)).

$$T_D = \max\{T_d : d = 0, \dots, N_c - 1\} \tag{15}$$

with N_c is the number of committed nodes.

Once the node D is selected, a vigilance criterion is evaluated over the input vector, the generator $\mathbf{I}_l$. Vigilance criterion causes the network to choose another committed node, if the one chosen does not comply with it. Mathematically it is represented by

$$\frac{|\mathbf{I}_l \,{}^\wedge \mathbf{w}_D|}{|\mathbf{I}_l|} \geq \rho \tag{16}$$

where ρ is the vigilance parameter, which lies in the interval [0,1].

If this criterion is respected, the network enters in resonance and the input vector, $\mathbf{I}_l$, is learnt according to

$$\mathbf{w}_D^{new} = \beta(\mathbf{I}_l \,{}^\wedge \mathbf{w}_D^{old}) + (1 - \beta)\mathbf{w}_D^{old} \tag{17}$$

where D is the index of the winning node and $\beta \in [0, 1]$ is the learning rate.

Otherwise a reset occurs, where a total inhibition is performed in the F_2 selected node, setting $T_D = 0$, and a new cycle search initiates. It is not needed to check the vigilance criterion for the other orientational inputs, $\mathbf{I}_n$ with $n \neq I$, because the one maximising the F_2 activity function, T_d is the one maximising the vigilance criterion ($\mathbf{I}_l$). If there is no committed node resonating the network, a F_2 non-committed node will be selected as the winner, so the winner will be $d = N_c$ and $N_c = N_c+1$.

5.2.2. Classification process

Once the network has been trained, it can be used as a classifier. The classification process is similar to the learning one, excepting there is no weight modification, the network learning is temporaly disabled $(\beta = 0)$, i.e. $\mathbf{w}_D^{new} = \mathbf{w}_D^{old}$.

An input vector is presented until the network enters in resonance. The texture label from that resonance node apprises us about the predicted texture, that is, the texture class for the input vector will be l with $L_{dl} = 1$ (d is the resonance node).

6. Tests and results

In order to show the processing nature of the depicted model, a comparison with other texture recognition methods is done. ChromARTex is evaluated under conditions as similar as those [7,12] used to compare to. These models perform good classifications of black and white Brodatz textures. Grossberg and Williamson's model is chosen to compare to since both models share the theoretical base. This comparative is intended to evaluate first accuracy and success in prediction, and then colour importance in texture classification.

The hybrid system's log-Gabor pyramid in [7] uses three levels, or spatial scales, and four orientations at each scale. Each level, after the first one, of the Gaussian pyramid is obtained by blurring the previous lower level (i.e. smaller spatial scale) with a Gaussian kernel (with standard deviation $\sigma = 1$) and then decimating the image (i.e. removing three out of four pixels in each 2×2 pixel block). This feature extraction scheme produce 15-dimensional feature vector for unsupervised learning and supervised labelling. In the labelling process, three methods are used: ITRULE

algorithm, K-nearest neighbour classifier and multilayer perceptron. From these, the hybrid system's log-Gabor with multilayer perceptron (HSG-MLP) achieved best results for a Brodatz textures test, so this will be the one chosen to compare with our model.

Grossberg and Williamson [12] proposed a self-organising neural system for recognising texture scenes, called ARTEX. It comprises an "ad hoc" BCS/FCS architecture, which generates textural patterns for a Gaussian ARTMAP classifier. Pre-processing stage yield a 17-dimensional recognition pattern composed of 16 textural features obtained from the BCS stage through Gabor filters using four scales and four orientations, and a filled-in brightness feature (FCS stage).

As explained in [12], to achieve same conditions it is necessary to procure same resolutions via blurring and decimation. This amount of blurring is equivalent to convolving with a single Gaussian kernel (6) with $\sigma_{resolution}$ ($\sigma = \sqrt{21} = \sqrt{1^2 + 2^2 + 4^2}$), which produces an 8×8 pixel resolution. That is, each patch of 8×8 pixels in the input image yields a single pixel in an output image for each oriented contrast feature. Analogously it is done for 16×16 and 32×32 resolutions. We include this resolution smooth in our model through

$$u_{ijk}^{(1,3)g} = t_{ijk}^{(1,3)g} \otimes G^{resolution} \tag{18}$$

where $t_{ijk}^{(1,3)g}$ is the smoothed output of the simple cells in position (i,j), orientation k and $g = 0,1,2$ according to the scale (see Eq. (9)); $G^{resolution}$ is a Gaussian kernel with $\sigma_{resolution}$ deviation.

We use four orientations as in [7,12] and three different scales, so the input pattern of the neural stage is a 26-dimensional feature vector (24 components from the simple cell stage output with scale- and resolution-related blurring processes applied and two components from the diffusion stage):

$$\mathbf{Q}_{ij} = [u_{ij0}^{(1)0}, u_{ij0}^{(1)1}, u_{ij0}^{(1)2}, u_{ij0}^{(3)0}, u_{ij0}^{(3)1}, u_{ij0}^{(3)2},$$
$$u_{ij1}^{(1)0}, u_{ij1}^{(1)1}, u_{ij1}^{(1)2}, u_{ij1}^{(3)0}, u_{ij1}^{(3)1}, u_{ij1}^{(3)2},$$
$$u_{ij2}^{(1)0}, u_{ij2}^{(1)1}, u_{ij2}^{(1)2}, u_{ij2}^{(3)0}, u_{ij2}^{(3)1}, u_{ij2}^{(3)2},$$
$$u_{ij3}^{(1)0}, u_{ij3}^{(1)1}, u_{ij3}^{(1)2}, u_{ij3}^{(3)0}, u_{ij3}^{(3)1}, u_{ij3}^{(3)2},$$
$$V_{ij}^{L-M}, V_{ij}^{S+(L-M)}]$$

where subscript $n = 0,1,2,3$ for angles $k = 0°,45°,90°,135°$, and using the tree spatial scales ($g = 0,1,2$).

Next, we present the tests performed. Values presented along the text are used for all the test simulations.

6.1. Two-texture problem

We begin with a first simple test, the "two-texture problem", with which we intend to present the compatibility of the results obtained with the human visual system, and to compare them with the segmentation methods known. The textures image (see Fig. 6a) is composed of two near-regular textures (weave and brick) which are widely used in texture benchmarks [7,12].

Fig. 6b–d, display the different stage outputs of the ChromAR-Tex model. Fig. 6b includes the contour map for large scale, Fig. 6c and d display the two outputs coming from the diffusion stages. Those outputs will constitute the last two components of the recognition patterns. Fig. 6e presents the results from the Canny extractor, using the cvCanny( ) function with 1, 100, and 70 as parameters; and Fig. 6f shows the output of the pyramidal segmentation using the cvPyrSegmentation( ) function with 30 000, 30 000, and 7 as parameters. cvCanny( ) and cvPyrSegmentation( ) are functions from the Open Computer Vision Library (SourceForge), OpenCv [17].

Comparing the results, it can be clearly observed that the proposed architecture behaves in a compatible manner with the human visual system. The presented model detects a texture boundary contour map with perceptual behaviour by extracting the illusory contour which marks off both textures. ChromARTex perceptually differentiates two textures through filling-in processes controlled by the illusory vertical contour. Those two comparative methods do not exhibit a concordance with the visual system, and so both extraction and segmentation obtain worse quality visual results.

Furthermore, a recognition test was run with the two-texture image. A smooth value of $\sigma_{smooth} = \sqrt{21}$ was chosen for the textural patterns, which corresponds to a 8×8 resolution, that is, each patch of 8×8 pixels in the input image yields a single pixel in an output image for each orientation [12]. The image was divided into lower and upper parts. The patterns from the upper half were used for the training process. The network was then
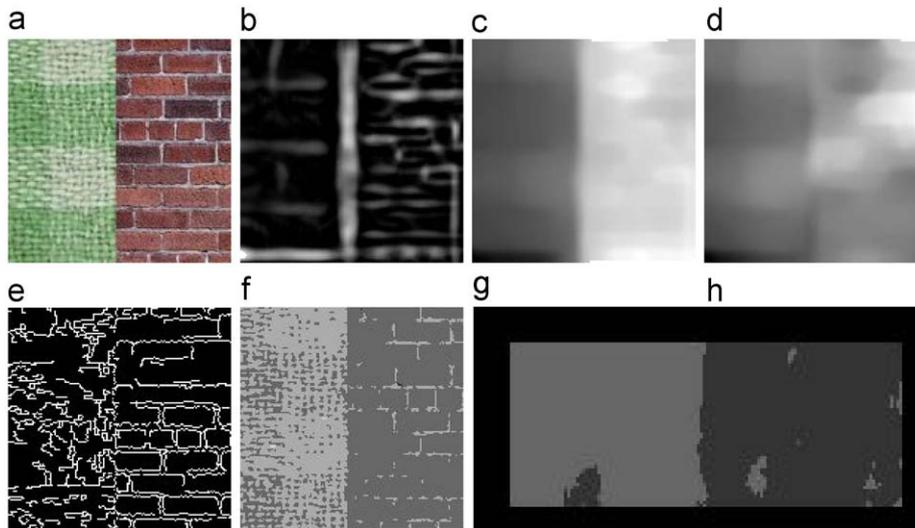

Fig. 6. Images of the two-texture test: (a) original image, 128×128 pixels, (b) image of the contours map for large scale, (c) output of the scale fusion stage for the $L-M$ channel, (d) output of the scale fusion stage for the $S-(L+M)$ channel, (e) image of extracted contours using Canny's extractor, (f) image segmentation with a pyramidal method, and (g) classification result of "two-texture" test. The darker grey level corresponds to the brick texture prediction while the lighter grey level corresponds to the weave texture prediction.

Fig. 7. 10-colour texture database (t1–t10, left to right). Textures extracted from VisTex database [28] (only one image per class is shown). Top row: structured textures. Bottom row: unstructured textures.

tested using the patterns coming from the lower half part. In the supervisory process, the categories created for the patterns on the left texture (weave) were associated to a class prediction pictured in light grey, while the patterns coming from the right texture (brick) were associated to another class prediction depicted in a darker grey.

In the training process as well as in the testing one, a frame of 10 pixels was left without any processing. In Fig. 6g we can see its resulting class prediction. The errors committed in the lower half prediction were of 115 points in the left side (weave texture) and 112 points in the right side (brick texture) which brings the error rate to a 3.17% (96.83% of success). Those statistics are of a similar magnitude to those obtained in [12], where a score of 95.7% was obtained for a texture mosaic test with 5 textures instead of two like in our case.

6.2. 10-Texture problem

In order to accomplish a further comparison of the different models for texture recognition, a test similar to the "10-texture library problem" proposed in [7,12] was run. For this test, we took 10 different texture classes from the VisTex colour-texture database (Massachusetts Institute of Technology) [28]. Every texture class has at least two associated images (see Fig. 7, only one image per class is presented). As Brodatz album just includes black and white texture images, for comparison purposes similar complexity colour texture images were chosen. Five images out of this set contained structured textures and the other five unstructured textures. Training was done with the first image in every texture class, using three different resolutions, as in [7,12], 8×8, 16×16 and 32×32. Testing process was performed over the second texture image.

Performance was measured configuring ChromARTex to use three scales, four orientations and one training epoch. Global detection rate are depicted in Table 1, 97.973% for 8×8, 98.343% for 16×16 and 98.561% for 32×32. The success rate rises for larger resolutions, even though detection rate is better in smaller resolutions for some textures. ARTEX model described in [12] obtains worst results for 8×8 and 16×16 resolutions (see Table 1), 95.8% for 8×8, 97.2% for 16×16, but a recognition rate of 100% for 32×32 (using four scales, four orientations and one training epoch). Best results for the hybrid system [7] are obtained with the MLP (multi-layer perceptron) but it needs many more training epochs (three scales, four orientations and 500 training epochs). The performance of this system is 94.5% for 8×8, 96.0% for 16×16 and 100% for 32×32 (see Table 1). ChromARTex obtained again best results with the two first resolutions.

Table 1

Statistics on 10-texture problem at three pixel resolutions, 8×8, 16×16 and 32×32, for HSG-ML, ARTEX and ChromARTex models.

Configuration	Class rate (%)	No. of epochs
8×8 *resolution*		
HSG-ML	94.5	500
ARTEX, all features	95.8	1
ARTEX, all features	96.3	5
ARTEX, 3 scales	97.1	5
ChromARTex	97.973	1
ChromARTex	98.799	5
16×16 *resolution*		
HSG-ML	96.0	500
ARTEX, all features	97.2	1
ChromARTex	98.343	1
32×32 *resolution*		
HSG-ML	100	500
ARTEX, all features	100	1
ChromARTex	98.561	1

Performance results for ARTEX and HSG-ML are obtained from [12] and [7], respectively.

ARTEX system provides its best results without using its largest scale, achieving 97.1% for 8×8 resolution (with 5 epochs and 300 samples/class). ChromARTex model achieves a better detection rate of 98.799% in similar conditions.

6.3. 30-Texture problem

ChromARTex was also tested for recognition of larger sets of textures. This section depicts the results achieved when working over a set of 30 different textures (see Fig. 8). Most of these textures where obtained from the colour texture images database VisTex [28]. Again, the tests were done using three scales and four orientations. For 16×16 and 32×32 pixel resolutions only one training epoch is used. For 8×8 resolution two different tests were done, the first one using 300 samples per class and 5 epochs and the second one using 1000 samples per class and only one training epoch. The aim of these configurations is testing the system in equivalent conditions to those used in [12].

As in the 10-texture problem, the results obtained using our model are better than those obtained by HSG-ML and ARTEX in the first two resolutions. The difference is greater for 8×8 resolution, no matter if using 300 samples per class and 5 epochs configuration or 1000 samples per class with one epoch configuration (Table 2).

Fig. 8. 30-colour texture set. Most of the textures extracted from VisTex database [28] (only one image per texture class is shown).

Table 2
Statistics on 30-texture problem at three pixel resolutions, 8×8, 16×16 and 32×32, for HSG-ML, ARTEX and ChromARTex models.

Configuration	Class rate (%)	No. of epochs
8 × 8 resolution		
HSG-ML	89.6	500
ARTEX, 300 samples/class	92.5	5
ARTEX, 768 samples/class	94.3	2
ChromARTex, 300 samples/class	96.515	5
ChromARTex, 1000 samples/class	97.4513	1
16 × 16 resolution		
HSG-ML	93.4	500
ARTEX	95.5	1
ChromARTex	95.8807	1
32 × 32 resolution		
HSG-ML	98.2	500
ARTEX	98.9	1
ChromARTex	96.606	1

Performance results for ARTEX and HSG-ML are obtained from [12] and [7], respectively.

Using the same set of 30 textures, series of tests were performed in order to study the behaviour of our system when modifying the number of textures learned. Along these tests the number of textures was increased from 5 to 30 (see Fig. 8), including one newer subset of 5 textures on each training and recognition processes. This analysis is similar to the one shown in [12], so both models can be compared. Those tests were run under similar conditions described in [12]. For 8×8 resolution, 300 samples per class and 5 training epochs were used. The same test was repeated using 1000 samples per class and one training epoch in order to not exceed the number of samples per class used in [12] to train the network, around 1500 (300 samples per class with 5 epochs and 768 samples per class with 2 epochs). Fig. 9a shows average classification rates for five different executions. The detection rate decreases when the number of learned textures increases. Best results are achieved when tests are performed with 1000 samples per class instead of 300, just as in ARTEX results are better with 768 samples per class rather than 300. Fig. 9b shows the detection rate for the three resolutions. It can be observed that better classification rate is obtained for the lager resolution (32×32).

6.4. Multitex problem

ChromARTex was also tested, using same configuration, over a "multitex problem", analogous but more complex than the

"texture mosaic problem" proposed in [7,12]. Our mosaic includes nine textural areas versus their five textural areas. Texture classes were selected from the set used in the 30-texture problem. We have included the same texture class in sixth and seventh areas to test their transitions in the face of different textures and their classification rate.

As explained before, the first image from each texture class was used to build a 210×210 pixels multitex test image (see Fig. 10a) in order to evaluate the frontier precision between textures in the prediction of our architecture.

Fig. 10b–d show the segmentation outputs of the proposed model: the boundary contour map for large scale and the scale fusion of the chromatic diffusions (L–M and S–$(L+M)$). The last two images will constitute the last part of the input pattern to the recognition stage, helping decisively to discriminate the image regions.

The results are presented in Table 3. Those results show a better class rate in all resolutions than those obtained in [12] (from 95.7% down to 79.5%, degrading performance at lower resolutions). The texture predictions of the hybrid system shown [7] for the 8×8 resolution visually appear to be less accurate than those obtained by ChromARTex; its value does not come out in their paper.

The images of the predictions can be seen in Fig. 11. Each prediction class is depicted with a different grey level. Those images reveal several remarkable points. First, sixth and seventh areas are classified in the same prediction class, as they should do. Furthermore, transitions of both areas have obtained similar good results.

The best prediction for the interior points shows up for a 32×32 resolution. However, it is the 8×8 resolution the one which accurately resolve texture transitions.

For a correct recognition, an accurate classification is needed not only for the region interiors, but also for the region transitions. Here, in the latter problem, is where chromatic enhancement component is very useful. During diffusion, image regions surrounded by closed boundaries tend to obtain uniform chromatic features, as different as possible to its closest regions, so transitions are remarkably discerned. This feature can be appreciated in Fig. 10c and d.

Fig. 12, shows the prediction images, but without using the chromatic enhancement outputs (Fig. 10c and d), that is, using the simple cells output, after smoothing, as the input pattern to the recognition stage. Results obtained are: 88.01% correct at 8×8 resolution, 87.95% at 16×16, and 86.35% at 32×32. It can be observed that errors committed in transitions are quite important. Not only the global rates (Table 3) are better using the outputs of the scale fusion of the chromatic diffusions, but transition areas have undoubtedly improved.
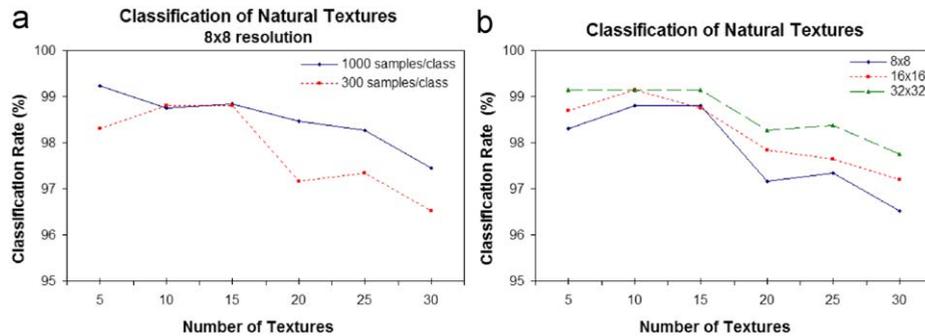
Fig. 9. Classification rate of ChromARTex: (a) for 8×8 resolution with different number of training samples (300 samples/class, 5 epochs and 1000 samples/class, 1 epoch), and (b) for the three resolutions used (8×8, 16×16, and 32×32).
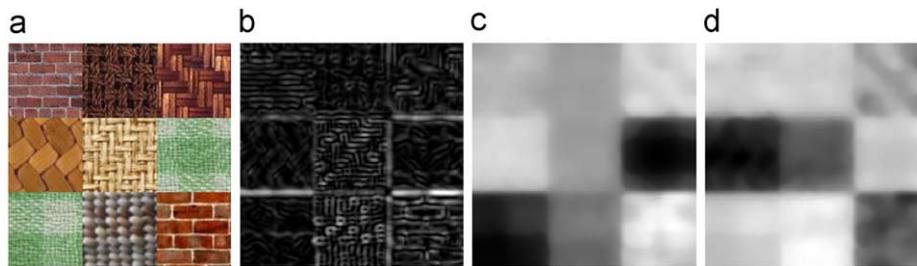

Fig. 10. (a) Multitex test image, 210×210 pixels, (b) image of the contours map for large scale, (c) output of the diffusion stage for the $L-M$ channel, and (d) output of the scale diffusion stage for the $S-(L+M)$ channel.

Table 3
Multitex prediction statistics.

Resolution	Class rate (%)
8×8	95.89
16×16	96.67
32×32	97.30

Although using the diffusion outputs has improved both global rate and transitions, it can be observed that region interiors are slightly deteriorated (Fig. 12 versus Fig. 11). It is not very noticeable, but it occurs in some textures. This is due to each region from the diffusion stage outputs not being entirely uniform, so interior rate decreases in comparison to what happens in Fig. 12.

To remark the utility of using the outputs of the diffusion stage to help the recognition, it is important to analyse the discriminability of the image regions based on their chromatic features. Fig. 13 shows the brightness distributions of the chromatic channels ($L-M$ and $S-(L-M)$) of four (t1, t2, t3 and t5) of the nine regions of the multitex test image from the output of the type 2 cells ((a) and (b)) as compared to the output of the diffusion stage ((c) and (d)). In the first case, the region distributions are widely overlapped while in the second case, each histogram is more clearly differentiated collecting most of the points in a narrower range, disjointing and so, differentiating more each of the distributions.

In summary, we can state that there are three differences between ChromARTex and the compared methods. First and most important one is the inclusion of two colour features versus a brightness feature. Second difference is that outputs from diffusion stage ($L-M$ and $S-(L+M)$), included in the recognition process, are inherent attributes of regions colour against contrast signal which depends on the surroundings of the region. Third one is the use of one additional receptive field in the pattern's textural components. Our architecture also includes in the patterns the processing of the symmetric receptive field with central excitation simple cells. Results obtained support that colour information ($L-M$ and $S-(L+M)$ opponent channels) and symmetric filtering (as asserted in [19,26]) contribute with important information for better classification and yield more accuracy to texture vision.

7. Conclusions

This work presents a multiple scale neural architecture for recognising coloured and textured images called ChromARTex. It has been compared to other recognition methods with tests and trials similar in difficulty to those presented by other authors in their works. The tests performed show the good features displayed by this architecture in order to process and recognise coloured textures.

Firstly, the "two textures test" has shown the compatibility in their responses between the depicted architecture and the human visual processing system and gets better marks than other quality proven methods like Canny's contour extraction and pyramidal segmentation method. When comparing it to Canny's contour extractor we have observed that the perceptual contour extraction used by our model displays features not present in Canny's like, for example, illusory contours detection and perceptual grouping which helps us to obtain results much more in line with those of human visual perception. ChromARTex has obtained visual results much more rewarding and much more similar to those of human perception than other segmentation processes, like pyramidal segmentation method. Besides, this test also shows that the
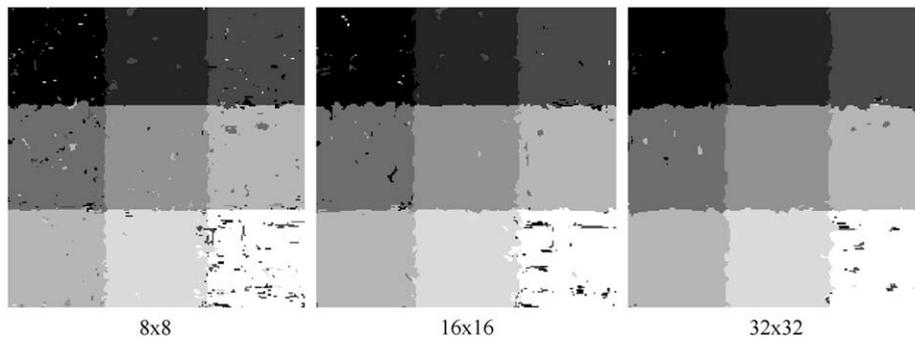
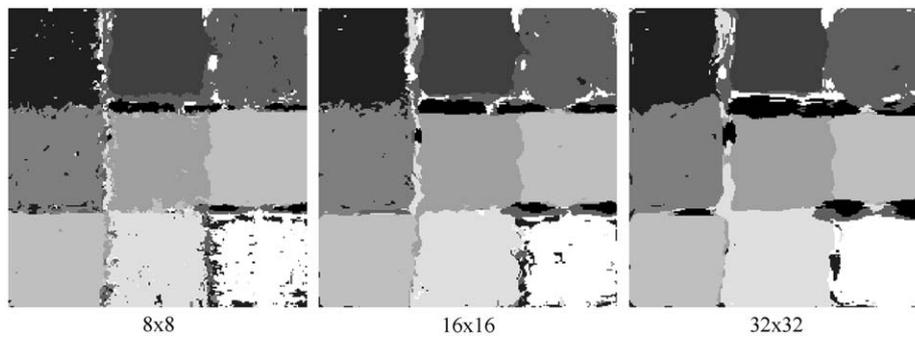**Fig. 11.** Multitex prediction results for three resolutions 8×8, 16×16, and 32×32.

Fig. 12. Multitex prediction results for three resolutions 8×8, 16×16, and 32×32 but without using the outputs of the diffusion stage ($L-M$, $S-(L+M)$). See Fig. 10c and d.
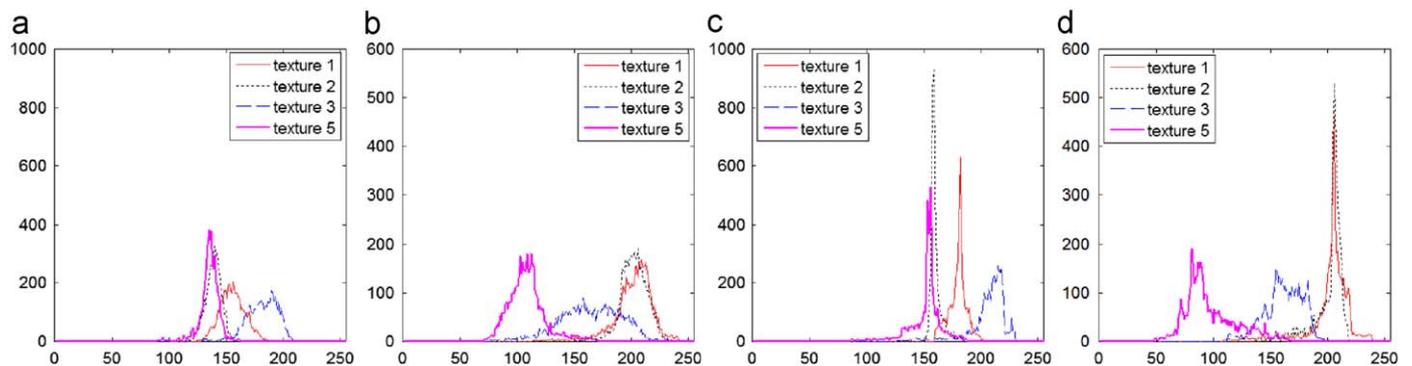

Fig. 13. Brightness distributions of the two chromatic channels ($L-M$ on (a) and (c) and $S-(L+M)$ on (b) and (d)) of four of the nine regions from the multitex test image. Only four regions (textures 1, 2, 3 and 5) are presented to obtain a clearer representation. (a) and (b): output of the type 2 cells (scale mean, to compare under similar conditions). (c) and (d): output of the scale fusion of the diffusion stage, where a differentiation of the distributions can be observed.

architecture obtains slightly better results when recognising than the architectures proposed in [7,12].

As a token of a deeper recognition comparative, three more test are proposed in order to measure the success percentage and the precision in the prediction of the boundaries of the textures. The tests were performed within the "10-textures problem", the "30-textures problem" and the "multitex problem" revealing the good results obtained by ChromARTex, getting better ratings than those included in [7,12].

Hence, hereby in this paper is presented not only a neural model to recognise colour texture images but also a reliable configuration to perform different tests, either with libraries of textures or with texture mosaics, achieving good results and demonstrating that the use of colour brings us better results in texture perception.

Acknowledgement

References

[1] G.A. Carpenter, Default ARTMAP, in: Proceedings of the IJCNN'03, Portland, Oregon, 2003, pp. 1396–1401.
[2] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps, Neural Networks 3 (5) (1992) 698–713.
[3] G.A. Carpenter, S. Martens, O.J. Ogas, Self–organizing information fusion and hierarchical knowledge discovery: a new framework using ARTMAP neural networks, Neural Networks 18 (2005) 287–295.

[4] S.E. Chelian, G.A. Carpenter, A neural model of colour vision, with applications to image processing and classification, Technical Report CAS/CNS TR-2005-006, Boston University, Boston, MA, 2005.

[5] J.G. Daugman, Two-dimensional spectral analysis of cortical receptive field profiles, Vision Research 20 (1980) 847–856.

[6] R.C. Gonzalez, R.E. Woods, Digital Image Processing 2/E, Prentice-Hall, New York, 2002.

[7] H. Greenspan, Non-parametric texture learning, in: S. Nayar, T. Poggio (Eds.), Early Visual Learning, Oxford University Press, London, 1996.

[8] S. Grossberg, Outline of a theory of brightness, colour, and form perception, in: E. Degreef, J. van Buggenhault (Eds.), Trends in Mathematical Psychology, North-Holland, Amsterdam, 1984.

[9] S. Grossberg, S. Hong, A neural model of surface perception: lightness, anchoring, and filling-in, Spatial Vision 19 (2) (2006) 263–321.

[10] S. Grossberg, E. Mingolla, J. Williamson, Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation, Neural Networks 8 (1995) 1005–1028.

[11] S. Grossberg, E. Mingolla, Neural dynamics of perceptual grouping: textures, boundaries, and emergent segmentations, in: S. Grossberg (Ed.), The Adaptive Brain II, North-Holland, Amsterdam, 1988 (Chapter 3).

[12] S. Grossberg, J.R. Williamson, A self-organizing neural system for leaning to recognize textured scenes, Vision Research 39 (1999) 1385–1406.

[13] M.A. Hoang, J. Geusebroek, A. Smeulders, Colour texture measurement and segmentation, Signal Processing 85 (2005) 265–275.

[14] S. Hong, S. Grossberg, A neuromorphic model for achromatic and chromatic surface representation of natural images, Neural Networks 17 (2004) 787–808.

[15] D.H. Hubel, Eye, Brain and Vision, Scientific American Library, vol. 22, W.H. Freeman, New York, 1995, p. 70.

[16] D.H. Hubel, M.S. Livingstone, Colour and contrast sensitivity in lateral geniculate body and primary visual cortex of the macaque monkey, The Journal of Neuroscience 10 (7) (1990) 2223–2237.

[17] SourceForge—Open Source Software, Open Computer Vision Library, 2008, 〈http://sourceforge.net/projects/opencvlibrary/〉.

[18] A. Jain, G. Healey, A multiscale representation including opponent colour features for texture recognition, IEEE Transactions on Image Processing 7 (1) (1998) 124–128.

[19] B. Julesz, J.R. Bergen, Textons, the fundamental elements in preattentive vision and perception of textures, in: O. Firschen (Ed.), Readings in Computer Vision: Issues, Problems, Principles, and paradigms, Morgan kaufmann, 1987.

[20] M.S. Landy, J.R. Bergen, Texture segregation and orientation gradient, Vision Research 31 (4) (1991) 679–693.

[21] M. Mirmehdi, M. Petrou, Segmentation of colour textures, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2) (2000) 142–159.

[22] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, colour, and texture cues, IEEE Transactions on Pattern Analysis and Machine Intelligence (2004) 530–549.

[23] E. Mingolla, W. Ross, S. Grossberg, A neural network for enhancing boundaries and surfaces in synthetic aperture radar images, Neural Networks 12 (1999) 499–511.

[24] G. Paschos, Perceptually uniform colour spaces for colour texture analysis: an empirical evaluation, IEEE Transactions on Image Processing 10 (6) (2001) 932–937.

[25] H. Permuter, J. Francos, I. Jermyn, A study of Gaussian mixture models of colour and texture features for image classification and segmentation, Pattern Recognition 39 (2006) 695–706.

[26] D.A. Pollen, S.F. Ronner, Visual cortical neurons as localized spatial frequency filters, IEEE Transactions on Systems, Man, and Cybernetics SMC-13 (15) (1983) 907–916.

[27] B. Thai, G. Healey, Modelling and classifying symmetries using a multiscale opponent colour representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (11) (1998) 1224–1235.

[28] VisTex database, Massachusetts Institute of Technology, available at 〈http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html〉.

[29] Y. Zhong, A.K. Jain, Object localization using colour, texture and shape, Pattern Recognition 33 (2000) 671–684.

[30] E. Zrenner, Neurophysiological Aspects of Colour Vision in Primates, Springer, Berlin, 1983.

Míriam Antón-Rodríguez was born in Zamora, Spain, in 1976. She received her M.S. and Ph.D. degrees in telecommunication engineering from the University of Valladolid, Spain, in 2003 and 2008, respectively.

Since 2004, she is an assistant professor in the School of Telecommunication Engineering and a researcher in the Industrial Telematics Group of the Department of Signal Theory, Communications and Telematics Engineering. Her teaching and research interests includes applications on the Web, bioinspired algorithms over intelligent transport systems, and neural networks for artificial vision.

Francisco Javier Díaz-Pernas was born in Burgos, Spain, in 1962. He received the Ph.D. degree in industrial engineering from Valladolid University, Valladolid, Spain, in 1993.

From 1988 to 1995, he joined the Department of System Engineering and Automatics, Valladolid University, Spain, where he has worked in artificial vision systems for industry applications as quality control for manufacturing. Since 1996 he has been a professor in the School of Telecommunication Engineering and a Senior Researcher in the Industrial Telematics Group of the Department of Signal Theory, Communications, and Telematics Engineering.

His main research interests are applications on the Web, intelligent transportation system, and neural networks for artificial vision.

José Fernando Díez-Higuera was born on 1963, in Valladolid, Spain. He graduated in industrial engineering in 1991 from the University of Valladolid, Spain. He received the Ph.D. degree in telecommunication engineering in 1996 from the latter university. From 1991 to 1999 he was an assistant professor at the Department of Signal Theory, Communications and Telematics Engineering of this University. Since 1999, he is a professor of Computer Programming in the same department.

His current research and teaching interests are in the areas of bioinspired algorithms, pattern recognition, neural networks and intelligent transport systems.

Mario Martínez-Zarzuela was born in Valladolid, Spain, in 1979. He received the M.S. degree in telecommunications engineering from the University of Valladolid, Spain, in 2004. Since 2005 he has been an assistant professor in the School of Telecommunication Engineering and a novel researcher in the Industrial Telematics Group of the Department of Signal Theory, Communications and Telematics Engineering.

His research interests include parallel processing on GPUs and neural networks development for artificial vision and object recognition.

David González-Ortega was born in 1972. He received his M.S. degree in telecommunications engineering from the University of Valladolid, Spain, in 2002. Since 2003 he has been a researcher in the Industrial Telematics Group of the Department of Signal Theory, Communications and Telematics Engineering. Since 2005, he has been an assistant professor in the Higher School of Telecommunication Engineering, University of Valladolid. His research interests include computer vision, image analysis, pattern recognition, neural networks and real-time applications.

Daniel Boto-Giralda was born in 1973. He received his M.S. degree in telecommunications engineering from the University of Valladolid, Spain, in 2000. Since 2004, he has been an assistant professor in the Higher School of Telecommunication Engineering, University of Valladolid and a researcher in the Industrial Telematics Group of the Department of Signal Theory, Communications and Telematics Engineering. His main teaching and research interests are real-time systems, biologically inspired optimisation algorithms and neural networks for intelligent transportation system and computer vision.