Pattern Recognition Letters 23 (2002) 1391–1398

Pattern Recognition Letters

www.elsevier.com/locate/patrec

Entropy-based representation of image information

Mario Ferraro [a], Giuseppe Boccignone [b,*], Terry Caelli [c]

[a] Dipartimento di Fisica Sperimentale, Università di Torino and INFM, via Giuria 1, 10125 Torino, Italy
[b] Dipartimento di Ingegneria dell'Informazione e Ingegneria Elettrica, Università di Salerno and INFM, via Melillo 1, 84084 Fisciano (SA), Italy
[c] Department of Computing Science, The University of Alberta, General Services Building, Edmonton, Canada, T6G 2H1

Abstract

Loss of information in images undergoing fine-to-coarse transformations is analysed by using an approach based on the theory of irreversible processes. In the case of grey level images, entropy variation along scales is used to characterize basic, low-level information and to identify perceptual components of the image, such as shape and texture. Here an extension of the approach to colour images is proposed. Spatio-chromatic information is defined, which depends on cross-interactions between the different colour channels. Examples illustrating the use of spatio-chromatic information are presented, related to pattern recognition and active vision.

Keywords: Scale space; Entropy; Color images; Feature encoding; Active vision

1. Introduction

A central issue in vision science is how the information contained in an image can be encoded in a way suited for the specific problem the visual system must solve. Typical methods to measure image information rely on the classical model (*a là* Shannon) which assumes the pattern itself as the message source. The model proposed here is different in that the message source is represented by a dynamical system, namely the pattern together with a given transformation. Images are considered as isolated thermodynamical systems, by identifying the image intensity with some dynamical variable, e.g. temperature or concentration of particles, evolving in time. According to thermodynamics (de Groot and Mazur, 1962), information in physical systems is related to the rate of variation of the entropy H, which is usually written as $\mathrm{d}H/\mathrm{d}t = \mathrm{d}H_e/\mathrm{d}t + \mathrm{d}H_i/\mathrm{d}t$, where $\mathrm{d}H_e/\mathrm{d}t$ is the variation due to external sources, while the entropy production term $\mathrm{d}H_i/\mathrm{d}t \geqslant 0$ is generated by changes inside the system; for the case of isolated systems, $\mathrm{d}H/\mathrm{d}t = \mathrm{d}H_i/\mathrm{d}t$. In general, $\mathrm{d}H_i/\mathrm{d}t$ can be defined by spatial integration of entropy production density σ, which is a function of both position and time (de Groot and Mazur, 1962).

It has been shown (Boccignone et al., 2001; Ferraro et al., 1999) that entropy production in fine to coarse transformations of images implicitly provides a measure of the information contained in the original image. Moreover, parts of the image with different information content can be identified according to different amount of density entropy

* Corresponding author.
E-mail address: boccig@unisa.it (G. Boccignone).

0167-8655/02/$ - see front matter
PII: S0167-8655(02)00099-5

production. Indeed this method can provide a preliminary step towards the inference of shape and texture.

In this paper we generalize to colour images our thermodynamical approach. It is well known that colour is an important cue in the recognition of natural patterns (Wandell, 1995), even though most computational vision models work exclusively in the luminance domain. We assume colour images to be isolated systems whose components (channels) are open subsystems interacting with each other through generalized thermodynamical forces. From this assumption a density of entropy production is derived, which takes into account interactions between different channels, and provides a measure of spatio-chromatic information.

Two applications of this approach will be presented. On the one hand, the behaviour of the density of entropy production along scales is exploited to define a suitable feature space by means of which parts of the image bearing different spatio-chromatic information content can be distinguished. On the other hand, spatio-chromatic information is used in an active vision context. In active vision, particular locations in the scene are selected based either on their behavioural relevance or on local image cues (Ballard, 1991). It will be shown that spatio-chromatic information can be used to obtain a conspicuity map, which can drive the dynamic selection of the "focus of attention" (FOA).

2. Background: information from irreversible transformations

Let Ω be a subset of R^2, (x, y) denote a point in Ω, and the scalar field $I : (x, y) \times t \rightarrow I(x, y, t)$ represent a gray-level image. The parameter t defines the scale of resolution at which the image is observed; small values of t correspond to fine scales, while large values correspond to coarse scales. A scale transformation is given by an operator T that takes the original image $I(\cdot, 0)$ to an image at a scale t, namely, $T_t : I(\cdot, 0) \rightarrow I(\cdot, t)$. We assume T to be a non-invertible or *irreversible* transformation which cannot be run backward across scale. Irreversibility of T ensures that the causality principle is satisfied, i.e. any feature at coarse level "controls" the possible features at a finer level of resolution—but the reverse need not be true. As a consequence of these assumptions every image point will, under the action of T, converge to a set of equilibrium points, denoted by the set $\mathscr{I}^*$, while preserving the total intensity (Ferraro et al., 1999). Let I^* denote the fixed point of the transformation, that is, $I^*(\cdot, 0) = I^*(\cdot, t), \forall t$, and let I^* be stable, that is $\lim_{t\to\infty} I(\cdot, t) = I^*(\cdot)$. Relevant information contained in an image I can be measured in terms of the conditional entropy between I and the corresponding fixed point I^* under the transformation T (Ferraro et al., 1999): $H(f|f^*) = -\int\int_\Omega f(x, y, t) \times \ln(f(x, y, t)/f^*(x, y))\, \mathrm{d}x\, \mathrm{d}y$. Here f and f^* are the normalized versions of I and I^*, respectively.

The "dynamic information" of the image, under the process T, is provided by the evolution of $H(f|f^*)$ across scales:

$$\frac{\partial}{\partial t} H(f|f^*) = \frac{\partial}{\partial t} H(f) + \frac{\partial}{\partial t} \int \int_\Omega f(x, y, t) \times \ln f^*(x, y)\, \mathrm{d}x\, \mathrm{d}y,$$

where $H(f) = -\int\int_\Omega f(x, y, t) \ln f(x, y, t)\, \mathrm{d}x\, \mathrm{d}y$ is the Boltzmann–Gibbs entropy. If f^* is constant over Ω, then $(\partial/\partial t)H(f|f^*) = (\partial/\partial t)H(f)$, and the entropy production (de Groot and Mazur, 1962), is $\mathscr{P} = (\partial/\partial t)H(f)$. For simplicity we have dropped the scale variable t in H and $\mathscr{P}$. Fine-to-coarse transformations can be modelled, by a diffusion or heat equation (Lindeberg and ter Haar Romeny, 1994): $(\partial f(x, y, t))/\partial t = \nabla^2 f(x, y, t)$. In this case, it has been shown (Ferraro et al., 1999) that $\mathscr{P} = \int\int_\Omega f(x, y, t)\sigma(x, y, t)\, \mathrm{d}x\, \mathrm{d}y$, where $\sigma(x, y, t) = (\nabla f(x, y, t) \cdot \nabla f(x, y, t))/f(x, y, t)^2$ is the *density of entropy production* in the thermodynamical sense (de Groot and Mazur, 1962). Since $\mathscr{P} \geqslant 0$, $H(f)$ is an increasing function of t. Note, also, that $\lim_{t\to\infty} \sigma = \lim_{t\to\infty} \mathscr{P} = 0$. In case of anisotropic transformations the derivation of the formula for the evolution of $H(f|f^*)$ is similar to the one outlined above (Boccignone et al., 2001). In general, during anisotropic diffusion, entropy production is contrasted by a term due to the selectivity of the anisotropic process; conse-

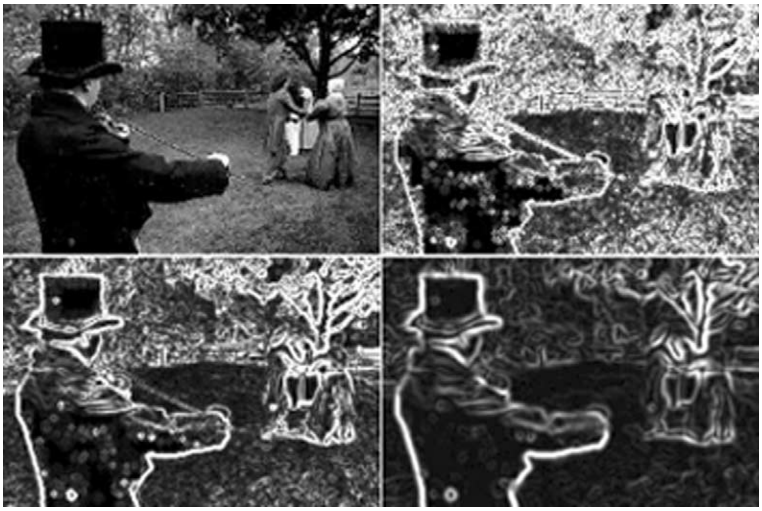

Fig. 1. The image "Old times" and σ maps computed at three different scales (from top to bottom right, 1, 5, 10 iterations of the diffusion process, respectively). Brighter points indicate high density of entropy production.

quently, $H(f|f^*)$ grows more slowly than in the isotropic case (Boccignone et al., 2001). Fig. 1 shows an example of local entropy production maps computed at different scales. For representation purposes the map $\sigma(\cdot, t)$ has been rendered as a grey level image, where brighter pixels represent higher σ values. It is easy to see that, during the transformation, σ is large along borders defining the shape of objects (e.g., the violinist in the foreground), smaller in textured regions (e.g., vegetation), and almost zero in regions of almost constant intensity.

In (Ferraro et al., 1999) it has been argued that $\mathscr{P}\,dt$ is the loss of information in the transition from scale t to $t + dt$. Intuitively, $\mathscr{P}$ is a global measure of the rate at which the image, under the action of T, looses structure. Analogously, the density of entropy production σ, being a function of x, y, measures the local loss of information (Ferraro et al., 1999). Thus, the overall behaviour of σ along scales depends on the interactions between neighbor diffusions which characterize local features of the image. This property has led to a method of feature identification (Ferraro et al., 1999), based on the notion of activity, which in the isotropic case is defined as $a_\sigma(x, y) = \int_0^\infty \sigma(x, y, t)\,dt$; similar definition can be given *mutatis mutandis*, for anisotropic diffusion (Boccignone et al., 2001). A thresholding procedure, enables the classification of every pixel in the image as belonging to one of three classes: low information or l-type pixel (low activity), medium information or m-type (medium activity), and high information or h-type (high activity). Identification of such basic types represents a preliminary step, both in the isotropic and anisotropic case, towards the labelling of image regions in terms of basic perceptual components: smooth regions, textures and edge/forms. Experiments have shown Boccignone et al. (2001) and Ferraro et al. (1999) that such a procedure entails the separation of well defined shapes characterized by high activity, from textured regions (Fig. 2). The use of anisotropic diffusion makes localization of different features in the image more precise. In particular, textural parts are more neatly encoded through m-type pixels, whereas localization of shapes, via h-type pixels, is improved, since anisotropic diffusion avoids edge blurring and displacement.

The activity function thus represents a preliminary way to encode parts of the images characterized by different information content; in the sequel, a different method will be shown, also based on local entropy production.

Fig. 2. Results of region identification by activity for image "Old times" obtained with the anisotropic process (20 iterations). Left image: h-type regions, right image: m-type regions.

3. Deriving spatio-chromatic information

To provide an extension of this framework to colour images it is necessary to address the issue of multi-valued isotropic diffusion. A colour image can be considered a vector-valued image, which we denote $\vec{f}(\cdot,t) = (f_i(\cdot,t))^{\mathrm{T}}$, where $i = 1,2,3$, labels the colour components, or channels, then a fine-to-coarse transformation must be defined for the vector field $\vec{f}(\cdot,t)$. Little work has been devoted to generalize isotropic diffusion to colour images, while some do address anisotropic diffusion of vector-valued images (Whitaker and Gerig, 1994; Sapiro and Ringach, 1996). The problem is commonly tackled by representing a colour image as a system of independent single-valued diffusion processes evolving simultaneously:

$$\frac{\partial f_i(x,y,t)}{\partial t} = \nabla^2 f_i(x,y,t). \tag{1}$$

Unfortunately, Eq. (1) does not model interactions among different colour channels. In contrast, there is a general agreement that the processing and interpretation of colour images must account to some extent for cross-effects, whatever the channels employed, RGB, HSI, etc. (Wandell, 1995). Indeed, cross-effects are important in complex images where interactions occur among colour, shape and texture (cfr. Fig. 3). Therefore, our basic assumption states that any colour image is an isolated system whose colour components are open subsystems interacting with each other through generalized thermodynamical forces. It is well known that the diffusion equation can be derived from a more general equation, namely $\partial f/\partial t = -\mathrm{div}\vec{J}$, where $\vec{J}$ is the flux density or flow (de Groot and Mazur, 1962). Also, it is known that, in a large class of transformations, irreversible fluxes are linear functions of the thermodynamical forces expressed by the phenomenological laws of irreversible processes. For instance, Fourier's law states that the components of the heat flow are linearly related to the gradient of the temperature. In this linear region then $\vec{J} = L\vec{X}$ (de Groot and Mazur, 1962), where $\vec{X}$ is called the generalized force and L is a matrix of symmetric coefficients L_{ij}.

Fig. 3. The "Waterworld" image (color, in the original).

We define, for each colour channel, the transition from fine to coarse scale through the equation $\partial f_i/\partial t = -\mathrm{div}\vec{J}_i$; for each i, the flow density is given by $\vec{J}_i = \sum_{j=1}^{n} L_{ij}\vec{X}_j$. We have chosen to model interactions among colour components setting $\vec{X}_i = \nabla(1/f_i(x,y,t))$ and $L_{ij} = \kappa_{ij} f_i f_j$, where κ_{ij} are coefficients weighting the strength of the interactions between channels i and j. For consistency with the theory of irreversible transformations the coefficients k_{ij} are assumed to be symmetric, $k_{ij} = k_{ji}$ (de Groot and Mazur, 1962). We then obtain the following system of coupled evolution equations:

$$\frac{\partial f_i}{\partial t} = -\mathrm{div}\left(\sum_j L_{ij}\vec{X}_j\right) = \sum_j \nabla \cdot \left(\kappa_{ij} f_i f_j \frac{\nabla f_j}{f_j^2}\right) \tag{2}$$

Eq. (2) can be developed as:

$$\frac{\partial f_i}{\partial t} = \nabla^2 f_i + \sum_{j \neq i} \kappa_{ij} \left\{ \frac{f_i}{f_j} \nabla^2 f_j + \frac{1}{f_j^2} \left[f_j \left(\frac{\partial f_i}{\partial x} \frac{\partial f_j}{\partial x} + \frac{\partial f_i}{\partial y} \frac{\partial f_j}{\partial y} \right) - f_i \left(\frac{\partial f_j}{\partial x} \frac{\partial f_j}{\partial x} + \frac{\partial f_j}{\partial y} \frac{\partial f_j}{\partial y} \right) \right] \right\}. \tag{3}$$

Then, colour evolution across scales, in the different channels, comprises a purely diffusive term and a non-linear term that depends on the interactions among channels; if $\kappa_{ij} = \delta_{ij}$ (independent channels), Eq. (3) reduces to Eq. (1).

By using the phenomenological coefficients L_{ij} defined earlier, local entropy production $\Sigma(x, y, t) = \sum_{i,j} L_{ij}\vec{X}_i \cdot \vec{X}_j$ (de Groot and Mazur, 1962) results to be:

$$\Sigma = \sum_i \frac{\nabla f_i \cdot \nabla f_i}{f_i^2} + \sum_{i \neq j} \kappa_{ij} \frac{\nabla f_i \cdot \nabla f_j}{f_i f_j}. \tag{4}$$

The *spatio-chromatic density of entropy production* Σ is, not surprisingly, made up by two terms: $\sigma_i = (\nabla f_i \cdot \nabla f_i)/f_i^2$, the density of entropy production for every channel i, considered in isolation, and the cross terms $\kappa_{ij}(\nabla f_i \cdot \nabla f_j / f_i f_j)$ that accounts for the dependence of entropy production on interactions among channels. Obviously, if colour channels are considered isolated, $\Sigma = \sum_i \sigma_i$.

4. Using spatio-chromatic information: examples

We will provide two examples illustrating the use of spatio-chromatic information. In order to specifically discriminate between chromatic and achromatic information, the original RGB image is converted to opponent colour space as defined in (Wandell, 1995). The numerical values of weighting coefficient κ_{ij} are $\kappa_{ij} = 0.5 \cdot \delta_{ij} + 0.5$. The 0.5 value has been empirically chosen since providing a good trade-off between the independent channel diffusion and unconstrained interaction. Other choices could be adopted in some cases, to signify the dominance of one channel over another (e.g., the R–G over the B–Y, (Wandell, 1995)).

The first example refers to a feature matching problem. Here, to fully exploit information provided by Σ evolution, a vector of values of Σ at different scales has been considered, rather than gauging local information via a single scalar, integral measure (activity function). Namely, for a finite number $k = 1, 2, \ldots, K$ of iterations of Eq. (3), at each point (x, y) of the image, a feature vector $\vec{v} = [\Sigma(x, y, t_1), \Sigma(x, y, t_2), \ldots, \Sigma(x, y, t_K)]^{\mathrm{T}}$ has been defined, which captures the evolution of Σ at (x, y) across different scales. A clustering procedure is then applied to the vector $\vec{v}$, a cluster being a natural group of points close in feature space. We need to subdivide our feature space in just three hypersurfaces, since, as previously discussed in case of grey level images, we are interested in the partitioning of *l*-type features, as opposed to *h*-type and *m*-type features. This is obtained by using the well known C-means clustering algorithm (Duda and Hart, 1973). This procedure produces three clusters, each corresponding to a set of image pixels. For actual partitioning a minimum number of $K = 10$ iterations were used. The results presented in Fig. 4 show that most of the *h* and *m*-type points are encoded in the luminance channel; chromatic information contributes mostly to *m*-type points and little to *h*-type points. This kind of distribution of *h*- and *m*-type points is not characteristic of the "Waterworld" image. Indeed similar results have been obtained for the other images of our data set (50 real world natural images collected from various sources). The layered spatio-chromatic representation presented here could provide the basis for a subsequent segmentation module or a two-level tracking module, which detects most salient points on the *h*-type layer and refines its search on the *m*-type ones.

The second example shows how spatio-chromatic information can be exploited in an active vision context. Here the problem of attention selection is a major one, since it allows to concentrate the visual process on a circumscribed region

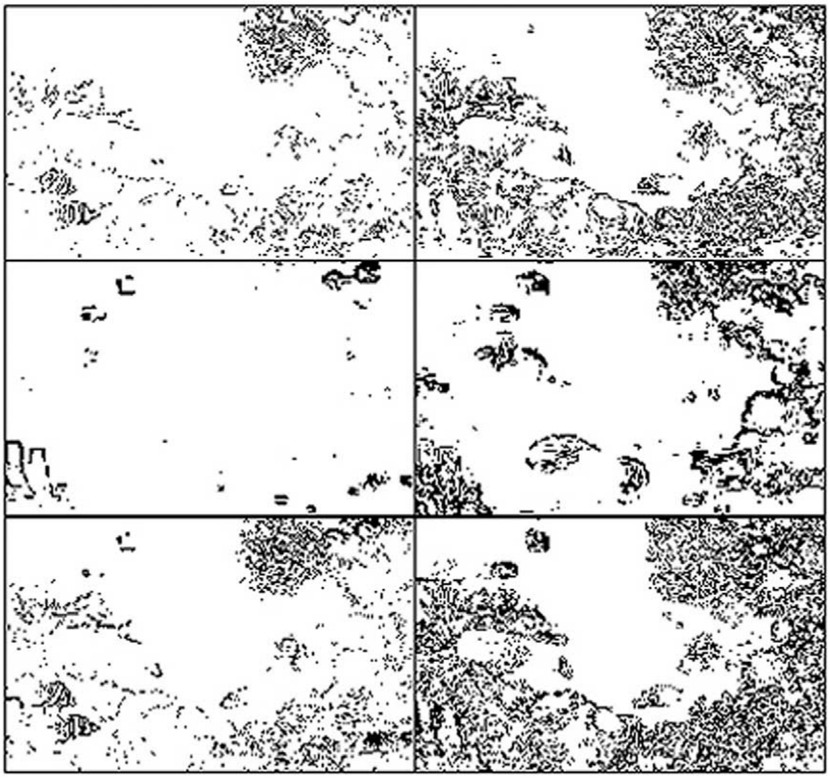

Fig. 4. Region identification for image "Waterworld". Left column: *h*-type regions; right column: *m*-type regions. First and second rows: results for independent luminance and color opponents channels, respectively; third row: result obtained taking into account cross-effects.

of the visual field, namely the FOA, which scans the scene either in a bottom-up, saliency driven fashion, and/or in a top-down, model driven way. In bottom-up processes, the attention mechanism computes some measure of relevance over the whole image and generates attentional shifts, to move the FOA on points of interest. It has been shown in the previous section that entropy Σ is an appropriate measure to encode parts of images characterized by different features. Here we use Σ to form a conspicuity map which represents the relevance of image points; for such purpose, an activity function is appropriate. For colour images we define activity as $a_\Sigma(x,y) = \int_0^\infty \Sigma(x,y,t)\,\mathrm{d}t$ and the map $a_\Sigma : (x,y) \rightarrow a_\Sigma(x,y)$ will be called an activity map. It is worth noting that since Σ is a rapidly decreasing function, a limited number of scales is needed to compute the map a_Σ. Next, a transformation $\mathscr{C} : a_\Sigma \rightarrow \mathscr{C}(a_\Sigma)$ is applied to a_Σ, which is implemented by the following steps: the values of a_Σ are normalized in a range $[0,S]$, so that $S = \max\{a_\Sigma(x,y)\}$; then, $\mathscr{C}(a_\Sigma(x,y)) = a_\Sigma(x,y)(S - \langle s_{a_\Sigma}\rangle)^2$, where $\langle s_{a_\Sigma}\rangle$ is the average of all local maxima; this transformation approximates a lateral inhibition mechanism (Itti et al., 1998). We call $\mathscr{C}(a_\Sigma)$ the *information-based conspicuity map*. Fig. 5 shows the $\mathscr{C}$ map of the "Waterworld" image obtained after 20 diffusion iterations. In turn $\mathscr{C}$ has been fed into a dynamical neural network, a 2D layer of leaky integrate-and-fire neurons representing a dynamical saliency map (DSM) as proposed by Itti et al. (1998). At any given time, the DSM evolution defines the most salient image location to which the FOA should be directed. Such network constitutes the input of a biologically inspired 2D "winner-take-all" (WTA) net-

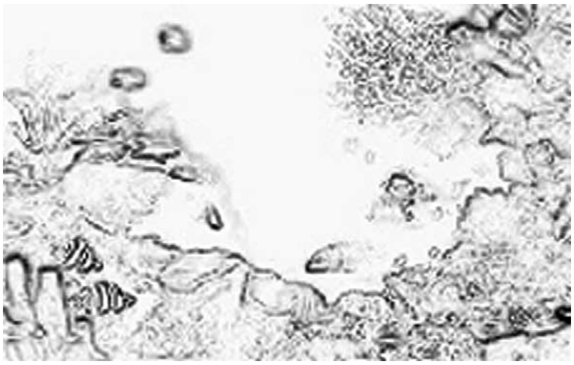

Fig. 5. $\mathscr{C}$ map of "Waterworld". Darker points represent higher values of $\mathscr{C}$.

work, which, on the basis of the DSM input, evolves until a neuron (the winner) fires; consequently, it shifts the FOA on the winning location, globally inhibits remaining WTA neurons, while triggering local lateral inhibition in the current FOA area of the DSM layer, so as to prevent immediate return to a just-attended location (see Itti et al. (1998) for details). In our simulation the evolution of both DSM and WTA networks continues until a certain number of FOA has been determined or the most part of the image has been analysed. Fig. 6 depicts the visual scanpath provided by the visual attention module on the image shown in Fig. 3 and on a region of interest which has been previously foveated in the same image. Our results are similar to those obtained by Itti et al. (1998).

5. Discussion and conclusion

In this paper a method for measuring local and global information in images has been generalized to deal with colour or vector-valued images. It has been shown that from local measures of variation of conditional entropy, a feature space can be defined so that each pixel is assigned to a partition characterized by different information content. One question here could be the choice of the dimension K of the feature space. This is related to a classical debate in the pattern recognition literature, concerning the number of features needed for a good classification. This problem in our case is overcome by the fact that, from a theoretical point of view, $\lim_{t\to\infty} \Sigma = 0$, and, practically, that a limited number of iterations (i.e. features), between 10 and 20, is necessary to encode most relevant information. Further, the method does not require to fix a specific scale for each type of activity class that is to be observed. It is also interesting to remark that h- and m-type classes are formed by taking advantage of full spatio-chromatic information, which is not simply the sum of the contributions of chromatic and achromatic channels.

The encoding through a feature vector can also be applied to gray level images. Experiments, not reported here due to space limitations, have shown that the clustering method provides a more accurate classification than the activity function, but at a higher computational cost. Thus, the choice between the two approaches should be driven by

Fig. 6. Left: FOA scanpath on "Waterworld". Right: A refined scanpath on a shape of interest extracted from the same image. Circles represent the foveation areas.

the desired trade-off between effectiveness and efficiency.

Finally, it has been shown that our approach can be effectively used in active vision. Here the spatio-chromatic information provides a suitable tool to compute a preliminary, bottom-up, conspicuity map to drive the attentive process towards the detection/recognition of shapes of interest.

References

Ballard, D., 1991. Animate vision. Artif. Intell. 48, 57–86.

Boccignone, G., Ferraro, M., Caelli, T., 2001. Encoding visual information using anisotropic transformations. IEEE Trans. Pattern Anal. Machine Intell. 23, 207–211.

de Groot, S.R., Mazur, P., 1962. Non-Equilibrium Thermodynamics. North-Holland, Amsterdam, The Netherlands.

Duda, R., Hart, P., 1973. Pattern Classification and Scene Analysis. Wiley, New York, N.Y.

Ferraro, M., Boccignone, G., Caelli, T., 1999. On the representation of image structures via scale-space entropy conditions. IEEE Trans. Pattern Anal. Machine Intell. 21, 1199–1203.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Machine Intell. 20, 1254–1259.

Lindeberg, T., ter Haar Romeny, B., 1994. Geometry-Driven Diffusion in Computer Vision. Kluwer, Dordrecht, The Netherlands.

Sapiro, G., Ringach, D., 1996. Anisotropic diffusion of multivalued images with application to color filtering. IEEE Trans. Image Proc. 5, 1582–1586.

Wandell, B., 1995. Foundations of Vision. Sinauer, Sunderland, Mass.

Whitaker, R., Gerig, G., 1994. Vector-valued diffusion. In: Lindeberg, T., ter Haar Romeny, B. (Eds.), Geometry Driven Diffusion in Computer Vision. Kluwer, Dordrecht, The Netherlands.