A Biologically Motivated and Computationally Tractable Model of Low and Mid-Level Vision Tasks

Iasonas Kokkinos[1], Rachid Deriche[2], Petros Maragos[1], and Olivier Faugeras[2]

[1] School of Electrical and Computer Engineering,
National Technical University of Athens, Greece
{jkokkin,maragos}@cs.ntua.gr
[2] INRIA, 2004 route des Lucioles, BP 93, 06902 Sophia-Antipolis Cedex, France
{Rachid.Deriche,Olivier.Faugeras}@sophia.inria.fr

Abstract. This paper presents a biologically motivated model for low and mid-level vision tasks and its interpretation in computer vision terms. Initially we briefly present the biologically plausible model of image segmentation developed by Stephen Grossberg and his collaborators during the last two decades, that has served as the backbone of many researchers' work. Subsequently we describe a novel version of this model with a simpler architecture but superior performance to the original system using nonlinear recurrent neural dynamics. This model integrates multi-scale contour, surface and saliency information in an efficient way, and results in smooth surfaces and thin edge maps, without posterior edge thinning or some sophisticated thresholding process. When applied to both synthetic and true images it gives satisfactory results, favorably comparable to those of classical computer vision algorithms. Analogies between the functions performed by this system and commonly used techniques for low- and mid-level computer vision tasks are presented. Further, by interpreting the network as minimizing a cost functional, links with the variational approach to computer vision are established.

1 Introduction

A considerable part of the research in the computer vision community has been devoted to the problems of low and mid level vision; even though these may seem to be a trivial task for humans, they are intrinsically difficult and the human visual system still outperforms by far the state-of-the-art. Therefore, its mechanisms could serve as a pool of ideas for computer vision research.

In this paper we propose a biologically motivated system for edge detection and image smoothing, apply it to real and synthetic systems, and explain it in computer vision terms. Our starting point is the system developed by Stephen Grossberg and his collaborators through a series of papers, which can be seen as the backbone for many researchers' models, e.g.[19,22]. While keeping the philosophy of the model intact, we propose using less processing stages and recurrent neural dynamics, resulting in a simpler yet more efficient model. Further, by

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3022, pp. 506–517, 2004.

building upon previous work on the connections between neural networks and variational techniques, we interpret the network model in variational terms.

Is section 2 we shall briefly present the architecture proposed by S. Grossberg and his collaborators for mid-level vision tasks, while in section 3 we shall propose our model, that uses a simpler and more efficient version of this architecture. In section 4 the interpretation of our model in computer vision terms is presented.

2 A Review of the Boundary/Feature Contour Systems

The model proposed by Stephen Grossberg and his collaborators in a series of papers [10,11,12,13,14], known as the FACADE (Form And Colour and DEpth) theory of vision is a versatile, biologically plausible model accounting for almost the whole of low- and mid-level vision, starting from edge detection and ending at segmentation and binocular vision (see references in [31]). Most of the ideas are intuitively appealing and relatively straightforward; however, as a whole, the system becomes complicated, in terms of both analysis and functionality. This is natural, though, for any model of something as complicated as our visual system; the model's ability to explain a plethora of psychophysical phenomena [11,31] in a unified way offers support for its plausibility and motivation for studying it in depth, trying to relate and compare it with computer vision techniques.

We have implemented the two essential components of the FACADE model, namely the Boundary Contour System (B.C.S.) and the Feature Contour System (F.C.S.). The B.C.S. detects and amplifies the coherent contours in the image, and sends their locations to the F.C.S.; subsequently, the F.C.S. diffuses its image-derived input apart from the areas where there is input from the B.C.S., signaling the existence of an edge; see Fig. 1(a) for a 'road map' of the system.

An Ordinary Differential Equation (ODE) commonly used by S. Grossberg to determine the inner state V of a neuron as a function of its excitatory (E) and inhibitory (I) input is:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -AV + (C - V)E - (D + V)I, \quad E = \sum_{n=1}^{N} w_n^e U_n,\ I = \sum_{n=1}^{N} w_n^i U_n \qquad (1)$$

where A is a passive decay term modeling the leakage conductance of a neuron; C and $-D$ are the maximal and minimal attainable values of V respectively; U_n are the outputs of neurons $1, \ldots, N$; $w_n^{\{e,i\}}$ are the excitatory/inhibitory synaptic weights between neuron n and the current neuron. The inner state V of the neuron is related to its output -a mean firing rate- U by a sigmoidal function; a reasonable choice is $U = G(V) = 1/(1 + \exp(-\beta(V - 1/2))$ so that for $V = 0$ we have a low output U. Equation (1) is closer to the neural mechanisms of excitation and inhibition than the common 'weighted sum of inputs' model and can account for divisive normalization and contrast invariance [3], which are helpful in visual processing tasks.

The Boundary Contour System

The outputs of cells in the B.C.S. are calculated as the rectified steady-state outputs of (1), so that all one needs to define are the excitatory and inhibitory

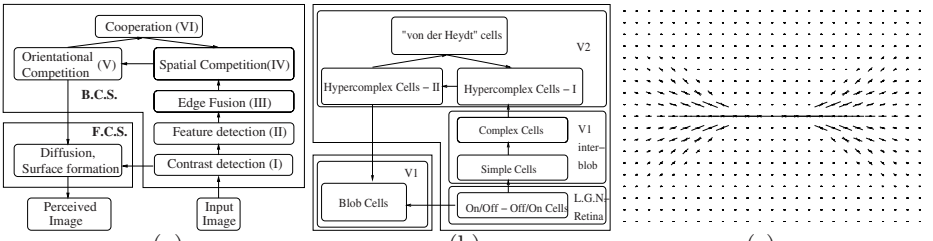


Fig. 1. (a) A block diagram of the B.C.S./F.C.S. architecture. (b) corresponding areas in the visual system (c) The shape of the lobes used for saliency detection in the horizontal direction, at stage VI. The 'needle' lengths are proportional to the weight assigned to each edge element in the corresponding direction and location

inputs; the processing stages used in the B.C.S. can be summarized as follows:

Stage I *Contrast Detection*: This stage models retinal cells, which exist in two varieties: one that responds to a bright spot surrounded by dark background, On/Off cells, and another that responds to the a dark spot on a bright background, Off/On cells. The excitatory input to an On/Off cell is modeled by convolving the input image F with a Gaussian filter of spread σ_1 and the inhibitory input is modeled by convolving with a Gaussian filter of spread $\sigma_2 < \sigma_1$; the roles of excitatory and inhibitory terms are swapped for Off/On cells.

Stage II *Elementary Feature Detection*: This is the function accomplished by simple cells in cortex area V1; their function is modeled by assuming each cell's excitatory/inhibitory input equals the convolution of the previous stage's outputs with a spatially offset and elongated Gaussian, with principal axis along their preferred orientation. Specific values used for the centers and spreads of this stage can be found in [31,14]. Given that there can be only positive cells responses, two varieties of cells are needed at this stage, one responding to increase and another to decrease in activation along their direction.

Stage III *Cue Integration*: At this stage, the outputs of simple cells responding to increases and decreases in image intensity are added to model complex cells, which are known to respond equally well to both changes. If color and/or depth information is used as well, this is where edge fusion should take place, based on biological evidence.

Stage IV *Spatial Competition*: The outputs of the previous stage are 'thinned' separately for each feature map, using Gaussian filters of different spreads to excite and inhibit each cell. The convolution of the complex cells' outputs with a Gaussian of large spread (resp. small spread) is used as an inhibitory (resp. excitatory) input. A novel term that comes into play is the feedback term from Stage VI, which drives the competition process towards globally salient contours; this results in a modified version of equation (1), where the constant feedback term is added to the competition process.

Stage V *Orientational Competition*: This time the competition takes place among neurons with the same position but different orientations; the goal is to derive an unambiguous edge map, where there can be only one tangent direction for a curve passing through a point. The excitatory input to a cell is the

activation of the Stage IV cell at the same location and orientation, while its inhibitory input comes from Stage IV cells with different orientations; the inhibitory weights depend on the angles between the orientations, becoming maximal for perpendicular orientations and minimal for almost parallel.

Stage VI *Edge Linking, Saliency Detection*: This is where illusory contours emerge; each neuron pools evidence from its neighborhood in favor of a curve passing through it with tangent parallel to its preferred orientation θ, using connections shown schematically in Fig. 1(c); mathematical expressions can be found in [14,31]. The output of a neuron at this stage is positive only when both of its lobes are activated, which prevents the extension of a line beyond its actual end. The output is sent to stage IV, thus resulting in a recurrent procedure, which detects and enhances smooth contours.

The Feature Contour System

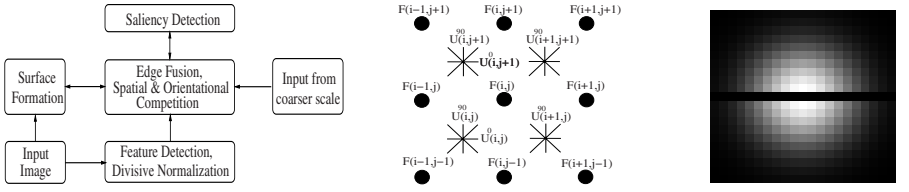
This is the complementary system to the B.C.S., where continuous brightness maps are formed, leading to the perception of surfaces. The image related input to this system is the output of the On/Off-Off/On cells of stage I; the formation of continuous percepts is accomplished by a process that diffuses the activities of the neurons, apart from locations where there is a B.C.S. signal, keeping the activities of neurons on the sides of an edge distinct. Specifically, the equations used to determine the activations of neurons $s(i,j)$ corresponding to On/Off or Off/On input signals, are:

$$\frac{\mathrm{d}}{\mathrm{d}t}s(i,j) = -As(i,j) + X(i,j) + \sum_{(p,q)\in N_{\{i,j\}}} (s(p,q) - s(i,j))P_{(p,q),(i,j)} \qquad (2)$$

where $X(i,j)$ is the steady-state output of Stage I (On/Off-Off/On cells) and $N_{\{i,j\}} = \{(i-1,j),(i+1,j),(i,j-1),(i,j+1)\}$. $P_{(p,q),(i,j)}$ is a decreasing function of the edge strength between pixels $(i,j),(p,q)$, which is computed using the steady state values of the B.C.S. stage V neurons. Physiological aspects of the mechanisms underlying the computation performed at this stage are discussed in [10], ch I, §24-26. Solving this system of ODEs with initial conditions the values of the On/Off- Off/On cells accomplishes the anisotropic diffusion of the On/Off cells activations, where the B.C.S. outputs determine where the diffusion becomes anisotropic; the perceived surfaces are modeled as the differences of the two steady-state solutions. According to the B.C.S./F.C.S. model, the perceived edges are at the points of intensity variation of the converged F.C.S. outputs.

3 A Simple and Efficient Biologically Plausible Model

In our implementations of the above stages we faced many problems: because so many parameters and interdependent stages are involved, the whole system becomes hard to tame; despite all our efforts, it was not possible to achieve a consistent behavior even for a limited variety of images. Since our first goal was to compare this system with some commonly used techniques for computer vision, and to see whether we can achieve an improved performance, we decided to simplify some of its stages, while keeping as close as possible to the original



(a) (b) (c)

Fig. 2. (a) The architecture of the new model (b) Relative positions of neurons and (c) Shape of lateral connections between neurons of the same orientation

architecture. Apart from the complexity and efficiency problems, most of our changes were inspired by the research experience accumulated among the computer vision community, which can be summarized as follows: (i) Edge thinning (stages IV-V) is an inherently *recurrent* process and not a feedforward one. The recurrent feedback term used in stage IV does not significantly help thin edges, even though it does help create new ones; recurrence should be used within each layer, using *lateral* connections among the neurons.(ii) The F.C.S. should be let to interact with the B.C.S., since the finally perceived edges are derived from the F.C.S.; therefore, it should play a role in the contour formation process, enhancing the more visible contours and exploiting region-based information. (iii) Results from coarse scales should be used to drive the edge detection process at finer scales. In the original B.C.S./F.C.S. architecture multiple scale results were simply added after convergence.

The model we propose, shown schematically in Fig. 2(a), is similar to the model of vision proposed by S. Grossberg, but is simpler and more efficient from a computational point of view. The architecture is similar to that proposed in [19], which, however, did not include a high level edge grouping process and a bottom up edge detection process, while the system was focused on texture segmentation. Our model performs both contour detection and image smoothing so it is different from other biologically plausible models like [22,20,16] that deal exclusively with boundary processing. We now describe in detail our model:

Stage I': *Feature Extraction, Normalization.* The first two stages (contrast & feature detection) of the B.C.S. are merged into one, using the biologically plausible Gabor filterbank [6] described in [19]; this filterbank includes zero-mean odd-symmetric filters while the parameters of the filters are chosen to comply with measurements of simple cell receptive fields. If we ignore the normalization and rectification introduced by (1), the cascade of the first two stages of the B.C.S. can be shown to be similar to filtering with a Gabor filterbank [31]. To account for divisive normalization, we used the mechanism of shunting inhibition, as in [3], where the feedforward input to each cell is determined from the beginning using a convolution with a Gabor filter, while the shunting term is dynamically changing, based on the activations of neighboring cells. The terms contributing to the shunting inhibition of neuron (i, j) are weighted with a Gaussian filter with spread equal to that of the Gabor filters; contributions from different orientations than θ are given equal weights, so that the equation driving the activity of a simple cell becomes:

$$\frac{\mathrm{d}V^{\theta,\sigma}_{\{o,e\}}}{\mathrm{d}t} = -AV^{\theta,\sigma}_{\{o,e\}} + (C - V^{\theta,\sigma}_{\{o,e\}})[F * \Psi^{\theta,\sigma}_{\{o,e\}}] - V^{\theta,\sigma}_{\{o,e\}} \sum_{\theta,o,e} G_\sigma * U^{\theta,\sigma}_{\{o,e\}} \qquad (3)$$

where $\Psi^{\theta,\sigma}_{\{o,e\}}$ are odd/even symmetric filters, with orientation preference θ at scale σ as in [19], while $U^{\theta,\sigma}_{\{o,e\}} = \max(V^{\theta,\sigma}_{\{o,e\}}, 0)$ are the time varying outputs of odd/even cells. For biological plausibility, we separate positive and negative simple cell responses, and perform the normalization process in parallel for each 'half', with the other half contributing to the shunting term; none of this appears in the above equations, for the sake of notational clarity.

Stage II': *Edge Thinning, Contour Formation.* At the next stage, an orientational energy-like term [1,25] is used as feedforward input:

$$E_{\theta,\sigma} = \sqrt{(U^{\theta,\sigma}_o)^2 + (U^{\theta,\sigma}_e)^2}\ . \qquad (4)$$

The difference lies in that simple cell outputs have already been normalized before, which results in a contrast invariant edge detector.

Edge sharpening in space and orientation is accomplished simultaneously with lateral connections among neurons of this stage. Specifically, the form of the lateral interactions among cells of the same orientation is modeled as the difference between an isotropic Gaussian and an elongated Gaussian whose longer axis is along the preferred orientation of the cells; this way fat edges are avoided, allowing a single neuron to be active in its vertical neighborhood, while collinear neurons support the formation of a contour. For example, for an horizontally oriented neuron located at $[i, j]$ the inhibitory connection strengths with neurons of the same orientation are (see also Fig.2(c)):

$$W^{0,0}_{[i,j]}[k,l] = \exp\left(-\left\{\frac{(i-k)^2 + (j-l)^2}{2\sigma_1^2}\right\}\right) - b\exp\left(-\left\{\frac{(i-k)^2}{2\sigma_1^2} + \frac{(j-l)^2}{2\sigma_2^2}\right\}\right) \qquad (5)$$

where $\sigma_1 > \sigma_2$. We use $W^{\theta,\phi}_{[i,j]}[k,l]$ to express the strength with which the neuron at $[k,l]$ with orientation ϕ inhibits the one at $[i,j]$ with orientation θ. The lateral inhibitory connections among complex cells of different orientational preferences are determined not only by the spatial but also by the orientational relationship between two neurons: almost collinear neurons inhibit each other, so that a single orientation dominates at each point, while perpendicular neurons do not interact, so that at corners the edge map does not break up. Such interactions among a neuron located at $[i, j]$ with horizontal orientation and a neuron located at $[k, l]$ with orientation θ can be expressed as:

$$W^{0,\theta}_{[i,j]}[k,l] = |\cos(\theta)| \exp\left(-\left\{\frac{(i-k)^2}{2\sigma_1^2} + \frac{(j-l)^2}{2\sigma_2^2}\right\}\right) \qquad (6)$$

Given that this time the competition process aims at cleaning the edge map, and not thinning it, we do not choose a specific direction for the Gaussian, so we set $\sigma_1 = \sigma_2$. Choosing this particular type of connection weights was based on simplicity, convenience and performance considerations [31].

Apart from the coarsest scale, the steady state outputs of the immediately coarser scale, $U^{\theta,\sigma+1}$, are used as a constant term that favors the creation of edges

at specific locations. The feedback term, denoted by $T(t)$, that is calculated at the following stage, is used throughout the evolution process, facilitating the timely integration of high-level information. Otherwise, if the system is let to converge before using the feedback term, it may be driven to a local minimum, so that it may be hard to drive it out of it. In addition to that, we coupled the evolution of the curve process with the evolution of the surface process, S; the magnitude of the directional derivative $|\nabla S_{\theta^\perp}|$ of the surface process perpendicular to θ helps the formation of sharp edges at places where the brightness gradient is highest. If the output of a neuron with potential $V^{\theta,\sigma}$ is $U^{\theta,\sigma} = g(V^{\theta,\sigma})$ the evolution equation for the activation of neuron $[i,j]$ at this stage is written as:

$$\frac{\mathrm{d}V^{\theta,\sigma}}{\mathrm{d}t} = -AV^{\theta,\sigma} + (C - V^{\theta,\sigma})I - (V^{\theta,\sigma} + D)\sum_{\theta'}\sum_{k,l} W^{\theta,\theta'}_{[i,j]}[k,l]U^{\theta',\sigma}[k,l] \quad (7)$$

$$\text{where} \qquad I(t) = [c_1 E + c_2 U^{\theta,\sigma+1} + c_3 T(t) + c_4 |\nabla S_{\theta^\perp}(t)|^2] \ .$$

The cues determining the excitatory input to the neuron, I, are bottom up - E, region based - $|\nabla S_{\theta^\perp}(t)|$, coarse scale -$U^{\theta,\sigma+1}$ and top-down -T. The weights, c_1, c_2, c_3, c_4, have been determined empirically [31] so as to strike a balance between the desire of being able to produce an edge in the absence of bottom-up input (E) and of keeping close to available bottom-up input.

Stage III': *Salience Computation.* This layer's outputs are calculated as the product of the two lobes' responses, which are continuously updated, resulting in a process that is parallel and continuously cooperating with Stage-II. Even though it is not clear *how* multiplication can be performed in a single cell, there is strong evidence in favor of multiplication being used in mechanisms as gain modulation; see also [22] for a neural 'implementation' of multiplication.

F.C.S. In our model surface formation interacts with boundary formation, keeping the boundaries close to places of high brightness gradient, thereby avoiding the occasional shifting of edges due to higher level (stage III') cues, or the breaking up of corners, due to orientational competition. The brightness values of the image were used, instead of the On/Off- Off/On outputs, in order to compare our algorithm to others. A less important architectural, but significant practical modification we introduced was that we considered the B.C.S. neurons as being located between F.C.S. neurons, as in [30], shown also in Fig. 2(b). This facilitates the exploitation of the oriented line-process neurons by blocking the diffusion among two pixels only when there is an edge (close to) *perpendicular* to the line joining them. More formally, this can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}S(i,j) = \sum_{\theta} \nabla_{\theta^\perp} S(1 - (U^\theta)) \quad (8)$$

where, as in [25] $S(i,j)$ is always the subtracted quantity in $\nabla_{\theta^\perp} S$. The equations we used were modified to account for the discrete nature of the neighborhoods and the relative positions of the F.C.S./B.C.S. nodes, but the idea used is the same: block diffusion only across edges and not along them.

Experimental Results and Model Evaluation

The results of our model are shown at the finest scale in Fig. 3; for all the images the same set of parameters has been used, while modifying these by 10% did not result in significant changes in the system's performance. We used a small fraction of the parameters used e.g. in [14]; we did not manage to achieve the same results using the original B.C.S./F.C.S. model , even when optimising its parameters for every single image; extensive results and implementation details can be found in [31]. It should be noted that our model operates at specific scales, so it may not respond in the same way to stimuli of arbitrary size.

Even though our model compares favorably to Canny edge detection on these test images, we do not claim that it is superior to the state-of-the-art [4,7,25, 27] in edge detection. We consider it more important that the modifications we introduced resulted in a simpler and efficient biologically plausible model, that does not largely deviate from the original B.C.S./F.C.S. architecture.

4 Interpreting the Model in Computer Vision Terms

The B.C.S./F.C.S architecture clearly parallels the usage of line and surface processes by computer vision researchers, see e.g. [2,9,30]; it is therefore of interest to interpret the previously described network in computer vision terms. We start with the line process-related functions of the network and continue with the function of the network as a whole.

Non-maximum Suppression: Non-maximum suppression is commonly used for the purpose of deriving a clean and coherent edge map from fuzzy inputs like the outputs of spatially extended filters. A common technique for nonmaximum suppression is to take the local maximum of the filter responses in the gradient direction and set the others to zero, like in [4,7,24]; in [2,9] a penalty term punishes spatial configurations of broad edges, while in [27] fitting the surface with a parabola and using the curvature and distance to the peak was proposed. Keeping to biological plausibility, our system implements non-maximum suppression by an analog winner-take-all type network (7), that continuously suppresses the activations of less active neurons, allowing the stronger ones to stand out (see [31] for details).

Perceptual Grouping and the Elastica Prior on Shapes: For the goal of enhancing perceptually salient contours various techniques have been developed by computer vision researchers; a similar pattern of pixel interactions as that shown in Fig. 1(c) is used in the voting technique of [15], in the probabilistic formulations used in [27,28] edge saliency is propagated among processing nodes, while in [26] a criterion including the squared integral of the curvature is used for edge linking. The popular and simple hysteresis thresholding technique used in [4] can be seen as performing some sort of edge grouping, as well as the penalty term for line endings used in [2,9].

In our model the contour linking process is cooperating with the contour formation process, driving the latter to the more salient edges, while avoiding using initial hard decisions from an edge detector output. The shape of the lobes used to perform perceptual grouping, shown in Fig. 1(c) was introduced

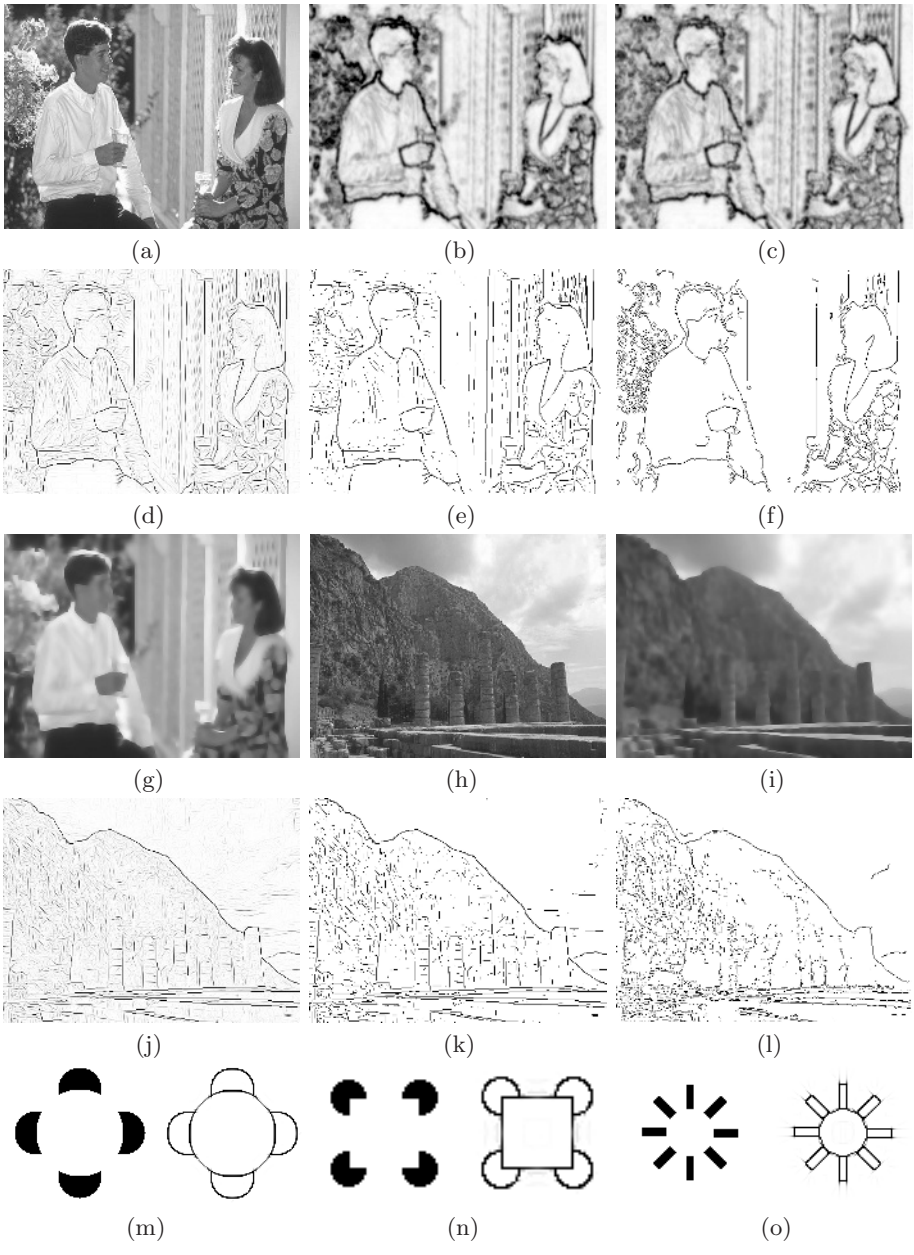


Fig. 3. Results of the proposed model: (a) input (b) orientation energy (c) normalized energy (d) line process at finest scale (e) thresholded results (f) Canny results with same number of pixels, $\sigma = 1, T_2 = .3, T_1 = .03$. (g) FCS outputs (h)-(l) same as (a),(g),(d)-(f) respectively, $\sigma = 3, T_2 = .1, T_1 = .01$ (m)-(o) results for 'Kanisza figures'.

by S. Grossberg and is now popular among researchers both in computer vision and biological vision [8,15,20,22,23]; this is natural since their shape, which favors low curvature contours, that occur frequently in our visual environment, enforces a reasonable prior on the contour formation process.

A very interesting link with the Elastica model of curves [21] is established in [28] where the shape of the lobes shown in Fig. 1(c) is related to the Elastica energy function $E(\Gamma) = \int_\Gamma 1 + a\kappa^2(s)\mathrm{d}s$, where κ is the curvature of the curve Γ. If a particle starts at $[0,0]$ with orientation 0 and the probability of its trajectory Γ is $P(\Gamma) \sim \exp(-E(\Gamma))$, then the probability $P(x,y,\theta)$ of the particle passing through x, y, θ is very similar to the lobes in Fig. 1(c). Loosely speaking, one could say that the high-level feedback term calculated using the lobes shown in Fig. 1(a) is related to the posterior probability of a contour passing through a point, conditioned on its surroundings, using the prior model of Elastica on curves.

A Variational Perspective: The link between recurrent networks and Lyapunov functions [5,17] has been exploited previously [18,19,29,30] to devise neural networks that could solve variational problems in computer vision; we take the other direction, searching for a variational interpretation of the model we propose. Even though based on [5] one can find a Lyapunov function for the recurrent network described in the previous section, the integrals become messy and do not help intuition; we therefore consider the simplified version of (7):

$$\frac{\mathrm{d}V^{\theta,\sigma}}{\mathrm{d}t}[i,j] = -AV^{\theta,\sigma} + CI - D\sum_{\phi}\sum_{k,l} W^{\theta,\phi}_{[i,j]}[k,l]U^{\phi,\sigma}[k,l] \tag{9}$$

where instead of the synaptic interaction among neurons we use the common sum-of-inputs model. The main difference is that the absence of the multiplicative terms in the evolution equations leads faster to sharp decisions and hence more probably to local minima. For simplicity of notation we drop the scale index, considering every scale separately and treat temporarily the excitatory input I as constant; a Lyapunov energy of the network is then [31]:

$$E = \sum_{i,j,\theta}[A\int_{1/2}^{U} g^{-1}(u)\mathrm{d}u - CIU] + D/2\sum_{i,j,\theta}\left[U^\theta[i,j]\sum_{\phi,k,l} W^{\theta,\phi}_{[i,j]}[k,l]U^\phi[k,l]\right] \tag{10}$$

In the above expression, $\int_{1/2}^{U} g^{-1}(u)\mathrm{d}u$ evaluates to $[U\ln(U) + (1-U)\ln(1-U)]/\beta + 1/2U - 1/4$ which consists of an entropy-like term, punishing 0/1 responses and $1/2U$ that generally punishes high responses; the first term is due to using a sigmoid transfer function and the second is due to shifting it by 1/2 to the right. The term $-IU$ lowers the cost of a high value of U, facilitating the emergence of an edge; among all U of fixed magnitude ($\sum_i U(i)^2$) the one that minimizes $-\sum_i U_i I_i$ is the one that minimizes $\sum_i (U-I)^2$, thus explaining the product terms in (10) as enforcing the closeness of U with I, without necessarily increasing U. The rightmost term in (10) can be expressed as [31]:

$$C(U) = \sum_{i,j,\theta}[G * U^\theta]^2 - b\sum_{i,j,\theta}[G^\theta * U^\theta(i,j)]^2 + \frac{1}{2}\sum_{i,j,\phi,\theta\neq\phi}\left[G^{\phi,\theta} * U^\theta\right]\left[G^{\theta,\phi} * U^\phi\right] .$$

The first two terms account for spatial sharpening of the edge map: G is an elongated Gaussian which is a scaled by $\sqrt{2}$ in space version of the first term in (5) and can be interpreted as disfavoring broad features like $G * U^\theta$. G^θ is an elongated Gaussian that is a scaled by $\sqrt{2}$ in space version of the second filter in (5), so this term guarantees that an isolated in both space and orientation edge will not get inhibited, due to the other two terms. $C(U)$ thus consists of both a diffusive term, namely a penalty on $G*U^\theta$ that wipes away broad structures and a reactive term $-G^\theta * U^\theta$ that acts in favor of the emergence of isolated edges. The last term accounts for orientational sharpening and is expressed in terms of scaled in space copies of the Gaussian filters used for orientational competition.

Putting everything together requires incorporating the interaction with the surface process; using equation (8) and assuming T^θ constant, the expression

$$E = \sum_{i,j,\theta} c_4(1-U^\theta(i,j))|\nabla_\theta S|^2 - U^\theta[c_1 I^\theta + c_2 T^\theta + c_3 U^\theta_{\sigma+1}] + A\int_{1/2}^{U} g^{-1}(u)\mathrm{d}u + C(U)$$

can be shown to be [31] a Lyapunov function of the system, since by differentiating w.r.t S and U^θ we get the evolution equations (8),(9) respectively. This functional can be seen as a more complex version of that introduced in [2], where a simple penalty term was used to enforce nonmaximum suppression and contour continuity to the anisotropic diffusion-derived line process.

5 Discussion

In this paper, motivated by ideas from biological vision, we proposed a simple and efficient model for low- and mid- level vision tasks which compares favorably to classical computer vision algorithms, using solely neural mechanisms. Its performance was demonstrated on both synthetical and real images, while its interpretation in computer vision terms has been presented, making the link with variational techniques. Extending our analysis to the whole FACADE model [11] of vision seems to be a promising future goal, which could result in a unified, biologically plausible model of computational mid-level vision.

References

1. E. Adelson, and J. Bergen. Spatiotemporal energy models for the perception of motion *JOSA*, 2(2), 284–299, 1985.
2. M. Black, G. Sapiro, D. Marrimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Trans. on Image Processing*, 7(3), 421–432, 1998.
3. M. Carandini and D. Heeger. Summation and division by neurons in visual cortex. *Science*, 264:1333–1336, 1994.
4. J. Canny, A computational approach to edge detection., *IEEE Trans. on PAMI*, 8(6):679–698, 1986.
5. M.A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Trans. on Systems, Man and Cybernetics*, 13(5):815–826, 1983.
6. J. Daugman. Uncertainty relation for resolution in space, spatial frequency and orientation optimized by two-dimensional visual cortical filters *JOSA*, 2(7):160-169, 1985.

7. R. Deriche Using Canny's criteria to derive a recursively implemented optimal edge detector *IJCV*, 1(2):167–187, 1987.
8. D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: Evidence for a local 'association field'. *Vision Research*, 33(2):173–193, 1993.
9. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the bayesian treatment of images. *IEEE Trans. on PAMI*, 66(6):721–741, 1984.
10. S. Grossberg. *Neural Networks and Natural Intelligence*. MIT Press, 1988.
11. S. Grossberg. 3-d vision and figure-ground separation by visual cortex. *Perception and Psychophysics*, 55:48–121, 1994.
12. S. Grossberg and E. Mingolla. Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perc. & Psych.s*, 38:141–171, 1985.
13. S. Grossberg and E. Mingolla. Neural dynamics of surface perception: Boundary webs, illuminants, and shape from shading. *CVGIP*, 37:116–165, 1987.
14. S. Grossberg, E. Mingolla, and J. Williamson. Synthetic aperture radar processing by a multiple scale neural system for boundary and surface representation. *Neural Networks*, 8(7-8):1005–1028, 1995.
15. G. Guy and G. Medioni. Inferring global perceptual contours from local features. *IJCV*, 20(1):113–133, 1996.
16. F. Heitger and R.v.d. Heydt. A computational model of neural contour processing: Figure-ground segregation and illusory contours. *proc. ICCV*, pp. 32–40, 1993.
17. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. of the N.A.S. of USA*, 81:3088 – 3092, 1984.
18. C. Koch, J. Marroquin, and A. Yuille. Analog neuronal networks in early vision. *MIT AI lab Technical report 751*, 1985.
19. T. S. Lee. A bayesian framework for understanding texture segmentation in the primary visual cortex. *Vision Research*, 35:2643–2657, 1995.
20. Z. Li. Visual segmentation by contextual influences via intracortical interactions in primary visual cortex, *Network:Comput.Neur.Syst.*, 10:187–212, 1999.
21. D. Mumford. Elastica and computer vision. In Bajaj J., editor, *Algebraic Geometry and its applications*, pages 507–518. Springer Verlag, 1993.
22. H. Neumann and W. Sepp. Recurrent V1-V2 interaction in early visual boundary processing. *Biological Cybernetics*, 91:425–444, 1999.
23. P. Parent and S.W. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. on PAMI*, 11(8):823–839, 1989.
24. P. Perona and J. Malik. Detecting and localizing edges composed of steps, peaks and roofs. In *Proc. ICCV*, pp. 52–57, December 1990.
25. P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. on PAMI*, 12(7):629–639, 1990.
26. A.Sha'ashua and S. Ullman. Structural saliency: the detection of globally salient structures using a locally connected network *Proc. ICCV*, pp. 321–327, 1988
27. X. Ren and J. Malik A probabilistic multi-scale model for contour completion based on image statistics. *Proc. ECCV*, pp. 312–327, 2002.
28. L.R. Williams and D.W. Jacobs. Stochastic completion fields: A neural model of illusory contour shape and salience. *Neural Computation*, 9(4):837–858, 1997.
29. A. Yuille. Energy functions for early vision and analog networks. *MIT A.I. lab technical report, 987*, 1987.
30. A. Yuille and D. Geiger A common framework for image segmentation. *IJCV*, 6:227–243, 1991.
31. I. Kokkinos, R. Deriche, O. Faugeras and P. Maragos Towards bridging the gap between biological and computational segmentation. *INRIA Research Report*, 2004.