Biol Cybern (2005) 93: 31–42
DOI 10.1007/s00422-005-0578-7

ORIGINAL PAPER

Biological Cybernetics

C. Rasche

Visual shape recognition with contour propagation

Received: 31 May 2004 / Accepted : 21 April 2005 / Published online: 8 June 2005

Abstract A neural architecture is presented that encodes the visual space inside and outside of a shape. The contours of a shape are propagated across an excitable neuronal map and fed through a set of orientation columns, thus creating a pattern which can be viewed as a vector field. This vector field is then burned as synaptic, directional connections into a propagation map, which will serve as a "shape map". The shape map identifies its own, preferred input when it is translated, deformed, scaled and fragmented, and discriminates other shapes very distinctively. Encoding visual space is much more efficient for shape recognition than determining contour geometry only.

1 Introduction

One may divide approaches to visual shape description into two classes, one pursuing a description by contour information, the other attempting to describe the region (or 2D space) that a shape engulfs. Contour-based approaches are abundant and come, for example, in the form of structural descriptions like part-based descriptions or neural networks like feature-integrating architectures (Palmer 1999; Rolls and Deco 2002). Region-based approaches in turn are rare: there exists the "filter" approach, which bears the idea to encode a visual image by some sort of Fourier transform (De Valois and De Valois 1988), but this approach has not pursued a specific model that attempts to encode shape in a neuronal network. An alternative region-based approach is contour propagation, an idea based on the Gestaltists' proposal to encode a shape by letting it "self-interact". We here follow this latter type because shape can be more distinctively represented from other shapes when region information is included.

Early Gestaltists like Koffka have already suggested that visual shape may be described using the shape self-interact by contour propagation (Koffka 1935). McCulloch had the idea that size-invariance of a shape could be encoded by concentric propagation of its contours (McCulloch 1965) (Fig. 1a), but he did not propose a specific model that would explain this type of shape representation. Blum was the first to have a specific model: the contour propagation process would find the so-called symmetric-axis of a shape, which are vectors that completely describe the space engulfed by a shape (Blum 1973) (Fig. 1b). We have already implemented this transform using a neuronal architecture (Rasche 2005b). However, the symmetric-axis transform only encodes the interior of a shape or contour form, but it does not encode the exterior, as for example, the space outside of a rectangle. We have considered extending Blum's symmetric-axis transform but that seems difficult to us. Rather we propose an architecture that remembers the flow of the outward- and inward-propagating contours as a whole (Fig. 1c): if one looks at the propagation direction of each contour point for each point in time, then one observes a pattern of vectors, which we now call the contour propagation field (CPF). This field is analogous to the optical flow field postulated in motion studies. If one obtains such a CPF for a specific shape and burns this CPF as synaptic weights into a propagation map, then this map will respond only to this shape and thus behave as what one could call a shape-sensitive (propagation) map. This idea is elaborated next by introducing the propagation map and by discussing modifications of it, which help to build an architecture that performs the envisioned recognition process.

(1) A propagation map is a layer of locally interconnected neurons that lets contours travel as a spiking wave. One can simulate such an excitable membrane using a meshwork of integrate-and-fire neurons connected by horizontal resistances, which allow for sub- and above-threshold spread of activity (Fig. 2a; Rasche 2005b). If the activity level in the map is raised across the spiking threshold then the input – the contour usually – will be actively prop-

C. Rasche
Department of Psychology, University of Notre Dame,
South Bend, IN, USA

C. Rasche
Department of Psychology,
417 Moore Building, Penn State, University Park, PA 16802, USA
Tel.: +814-8651930
Fax: +814-8637002
E-mail: cur12@psu.edu

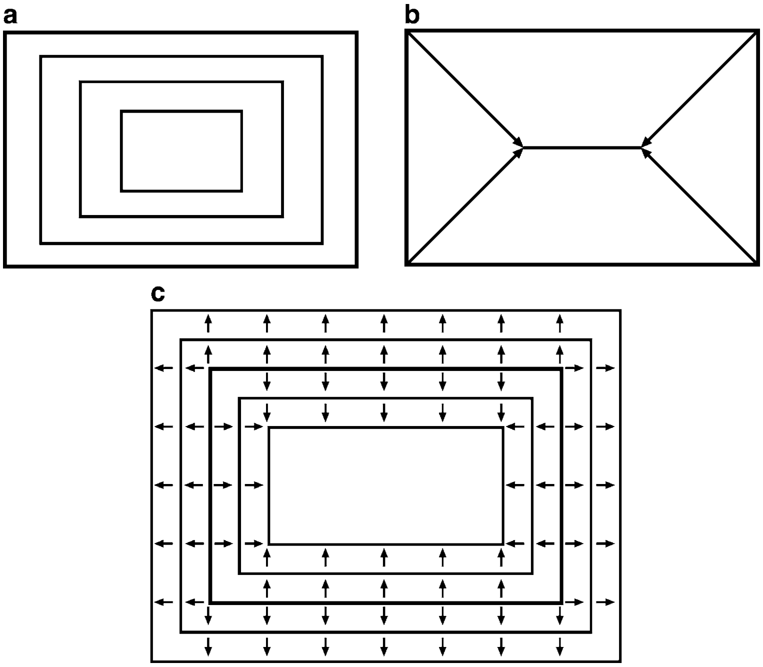

Fig. 1 Shape encoding techniques. **a** McCulloch's idea of concentric shape propagation to encode size-invariance (1965). **b** Blum's symmetrix-axis transform (1973). **c** The here proposed idea of encoding the contour-propagation field (CPF) – shape outlined in black, propagating contours outlined in gray: two 'virtual' time steps are shown; the vectors delineate the local propagation direction of the inward- and outward-propagating contour creating so an inward- and outward-pointing vector field

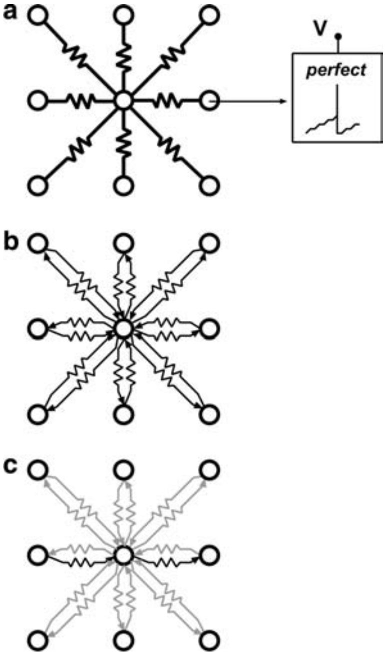

Fig. 2 Map connections. **a** Propagation map: integrate-and-fire neurons (*circle*) connected by horizontal resistors. **b** Same as a but with reciprocal, unidirectional connections. **c** Only one of the 'directed connections' is turned on (*black*), the others are turned off (*gray*): this map will prefer input moving toward the right only

agated across the membrane: for instance, a dot source triggers an annular outward-propagating wave, or a line triggers an oval-shaped outward-propagating wave.

(2) If the dynamics of such a propagation map are made inert, then the input will not be actively propagated but rather passively: it limitedly spreads through the subthreshold domain of the map. The parameters can be adjusted such that the map will only respond to continuous input – a motion input or – as in our case – a propagating contour: initial input will be firstly propagated and integrated in the subthreshold domain, but when the map's activity level has reached the spiking threshold, then the map starts spiking regularly – if the map continues to receive motion input. This map can be termed a motion-sensitive map.

(3) One can modify the motion-sensitive map such that it preferentially responds to a certain contour propagation (or motion input). To obtain this, firstly, the horizontal resistors are replaced by reciprocal but unidirectional resistors of equal strength (weight) (Fig. 2b). That alone would not change any of the propagation characteristics as described in point 2. But if one, secondly, started to tweak on the strength of the resistors, then only certain motion input patterns are preferred. For example, if all resistors are turned off but only the ones pointing toward the right are turned on (Fig. 2c), then only motion input toward the right leads to spiking in the motion-sensitive map. Thus, by selectively modifying the synaptic weights of the unidirectional resistors, the map can be made selective for certain contour-propagation input.

With these ideas, one can now put a shape-sensitive map together. A CPF can be obtained by continuously looking at the propagation dynamics of a shape through a set of orientation columns – like they exist in the primary visual cortex (Hubel and Wiesel 1968; Hubel 1995) – at each point in time. This shape-specific CPF is then burned into a map as described in (3): each vector turns on the corresponding unidirectional resistor in the map; all other synaptic connections at that location are turned off. This weight-burning action represents a one-shot learning process. After that, the map serves for recognition: it will preferentially respond to contour-propagating input whose shape is identical or similar: it will show a certain degree of size and translation invariance. Such a shape-sensitive map is hereafter called simply a shape map (SM).

We now describe an architecture and its simulation that performs the described processes. We then discuss in more detail why we pursue this type of region-based description.

2 Methods

Architecture The specific system that was developed is illustrated in Fig. 3. It consists of a propagation map (PM), a set of orientation columns (OCs), a set of direction-selective columns (DCs) and a set of SM, one for each shape. The graphical illustration describes the information flow within the architecture and the functionality of its components: it does not represent any specific anatomical findings; the term DCs merely expresses the use of a set of direction-selective cells, it does not suggest the actual existence of such columns in primary visual cortex.

Propagation map The PM behaves as described under point number (1) in the introduction and has been already described in (Rasche 2005b). We here only review the equations. It is a layer made of I&F neurons.

The neuronal voltage V_{PM} at location (x,y), at its next step, $t+1$, is given by its present potential plus the input of its neighboring neurons, $I_n(t)$, and initial external input I_e, the visual shape or other stimulation:

$$V_{PM}(x, y, t+1) = V_{PM}(x, y, t) + I_n(t) + I_e. \qquad (1)$$

I_n is the sum of positive membrane differences between the center neuron at location (x,y), V^c_{PM} and each of the eight neighboring neurons, V^k_{PM}, multiplied by the conductance, g_a:

$$I_n(t) = \sum_{k=1}^{8} \max[g_a(V^k_{PM}(t) - V^c_{PM}(t)), 0]. \qquad (2)$$

(The positive membrane difference is calculated for reasons of simplicity: it allows for an equally propagating spiking wave in an orthogonal grid. A possibly more realistic simulation is better done in a grid with hexagonal connectivity but that required major reprogramming of the following maps.) When V_{PM} exceeds the spiking threshold, V_{thres}, a spike of short duration is triggered, followed by a refractory period of short duration. The refractory period prohibits the bounce back of activity and ensures propagation into only one (local) direction. The activity of each PM neuron is binarized, $S_{PM}(x, y, t) = [\ 1 \text{ if } V_{PM}(x, y, t) > V_{\text{thres}}, 0 \text{ otherwise}]$. S_{PM} represents the set of spikes, which generally delineate the outward- and inward-propagating contours at each point in time.

Orientation Columns The propagating contours, S_{PM}, feed into a set of OCs responding to different 'wave' orientations running across the propagation map. A single OC consists of a set of cells whose receptive field (RF) consists of an excitatory and inhibitory part, taken from a 3 × 3 pixel field. The four 3 × 3 grids on the right side in Fig. 3 show the first three distinguishable orientations (0,22.5,45). The excitatory input is made of three sequentially aligned pixels (black pixels in grid), giving 24 aliased ('discretized') orientations, of which

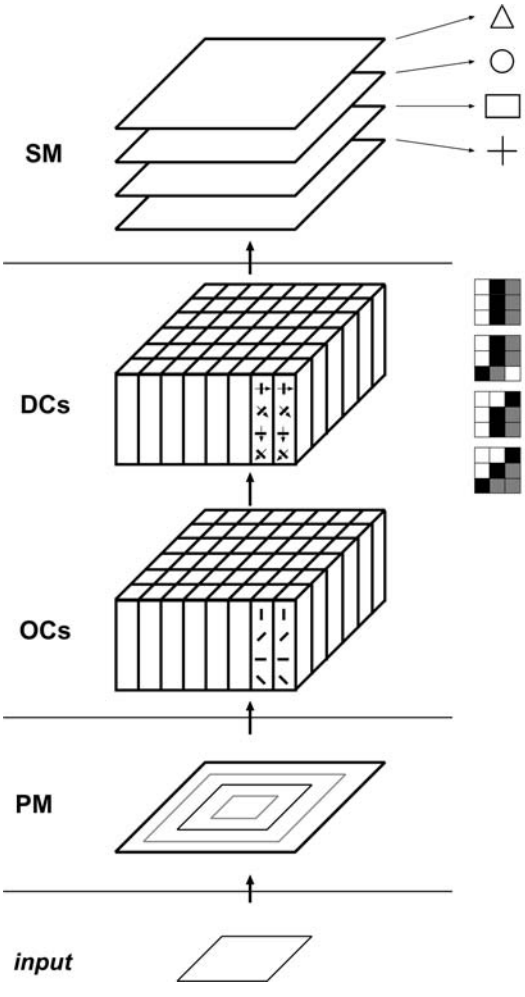

Fig. 3 Architecture. From bottom to top: the input shape (*input*) is dipped into the propagation map (*PM*), where it propagates inward and outward (*gray outlines*). The orientation columns (*OCs*) and the direction columns (*DCs*) determine the contour-propagation field (*CPF*) which is then fed into each shape map (*SM*). None of the components attempts to reflect any anatomical detail

only 16 are distinguishable (0, 22.5, 45, 67.5, 90, ... 337.5 degree). The inhibitory input (gray pixels in grid) consists of the pixels to one side of the excitatory pixels. The inhibitory input is necessary, because it ensures that only edges are signaled and not any cluster of excitation in the RF field, which may occur, for example, when the propagating contour wave is two pixels wide. Because of the use of this single strip of inhibitory inputs, the orientations range from 0 to 360 degrees and not only until 180 degrees.

An orientation-selective neuron (or OC neuron) is modeled as a leaky (or forgetful) I&F unit (Koch 1999). A leakage current is necessary in order to detect coincident input only. The voltage of an OC neuron at location (x, y) and orientation selectivity o is described as:

$$V_{OC}(x, y, o, t+1) = V_{OC}(x, y, o, t) + I_{RF}(t) - L, \qquad (3)$$

where L is a constant amount of leakage, $I_{RF}(t)$ is the RF input that is expressed as follows:

$$I_{RF}(t) = \sum_{k=1}^{3} S_{PM}(i_k, j_k, t) * A_e - S_{PM}(m_k, n_k, t) * A_i, \qquad (4)$$

with indices i, j corresponding to the three excitatory pixels, and indices m, n corresponding to the three inhibitory pixels, centered around the index (x,y) of the receiving OC neuron (x, y corresponds to the center of the 3×3 RF field). A_e and A_i are the respective synaptic amplitudes for the excitatory and inhibitory input. Their values are set such that an OC cell fires if all excitatory inputs are on and all inhibitory inputs are off simultaneously. When V_{OC} exceeds the spiking threshold, a spike is triggered, followed by resetting V_{OC} to 0. No refractory period is modeled in this unit. The voltage of OC neurons is binarized, $S_{OC} = \text{bin}(V_{OC})$, where bin is the threshold function as described for S_{PM} above. S_{OC} is then fed into the direction columns.

Direction Columns A piece of propagating contour will cause two OC cells to fire, one for each side of the contour wave, whose orientations are set apart by 180 degrees. Thus, the OC cells do not sense the direction of the propagating contour. One obtains this desired direction-selectivity by correlating two neighboring OC cells of the same orientation: if they have been activated sequentially into one direction, then that signals the direction of a piece of propagating contour. One could implement such direction selectivity in a single neuron by, for example, either using logic, synaptic interaction or by creating a direction-selective dendrite made of two compartments (Koch 1999). We have omitted such a biophysical simulation for reasons of simplicity and the correlation is performed in a mere 'bit-wise' manner. The correlation result is then placed into a set of DCs whose output can be expressed analogous to the OCs output: $S_{DC}(x, y, d, t)$. S_{DC} is then fed into each shape map. If this spiking output is observed over the entire time-course of contour propagation, then that 'flow-pattern' represents the CPF:

$$CPF = \sum_{t=1}^{\text{end}} S_{DC}(x, y, d, t). \qquad (5)$$

The number of directions remains the same as the number of orientations: there are 24 aliased directions of which only 16 are distinguishable (0, 22.5, 45, ... 337.5 deg).

Shape maps An SM is structurally and functionally basically the same as a propagation map, except that it possesses two unidirectional resistors between two neighboring neurons and that each resistor is individually adjustable in its strength (Fig. 2b, and c; see also point 3 in the introduction). The formalism is therefore very similar to the propagation map: the neuronal voltage V_{SM} at location (x, y), at its next step, $t + 1$, is given by its present potential, the input of its neighboring neurons, $I_n(t)$ and the input from the DCs, $I_{DC}(x, y, t)$:

$$V_{SM}(x, y, t+1) = V_{SM}(x, y, t) + I_n(t) + I_{DC}(x, y, t). \qquad (6)$$

I_n is the sum of positive membrane differences between the center neuron at location (x,y), V^c_{SM} and each of the eight neighboring neurons, V^k_{SM}, multiplied by a conductance of certain magnitude. The conductance pattern (or weight pattern) is expressed as $G(x, y, k)$, whereby k denotes the incoming connection that a neuron shares with a neighbor:

$$I_n(t) = \sum_{k=1}^{8} \max[G(x, y, k)(V^k_{SM}(t) - V^c_{SM}(t)), 0]. \qquad (7)$$

The input from the direction columns, I_{DC}, is the sum of activity of each direction column at each location (x, y):

$$I_{DC}(x, y, t) = \sum_{d=1}^{24} S_{DC}(x, y, d, t). \qquad (8)$$

The activity level of a shape map is also binarized, $S_{SM}(x, y, t) = bin(V_{SM}(x, y, t))$. The conductance $G(x, y, k)$ is generally adjusted lower than in the PM to make the SM dynamics inert to a certain degree (as discussed in point 2 in the introduction). The SM should only respond when it is continuously stimulated into the direction of the adjusted weights.

The architecture suffers from the problem of a coarse connectivity, in particular the SM. Specifically, a direction column provides 16 distinguishable directions, whereas an SM neuron has only eight neighbors and thus eight directions, which is too coarse for representing simple shapes distinctively. It would therefore make sense to explore PMs (or shape maps only) with different connectivity, but in our current simulation we stick to the current wiring for reason of convenience. The problem is therefore solved by using two propagation layers for each SM, SL_1 and SL_2, which are hereafter called shape layers in order to distinguish them from the term shape map. Each shape layer propagates the input from eight different directions. SL_1 bears the eight angles

0, 45, 90,..., 315 degrees, SL_2 contains the eight remaining angles 22.5,67.5,112.5,...,337.5 degrees. Hence, Eqs. 6, 7 and 8 are used twice, once for each layer, whereby in Eq. 8, the directions from the corresponding set of angles are used. The use of two shape layers, allows to distinguish between 16 directions.

Learning A novel shape is learned by placing it on the PM and obtaining its specific contour propagation field, CPF_s (Eq. 5). Learning is done with the centered shape – when it is placed into the middle of the PM (see also 'testing' below). This CPF_s then determines the connection strengths in a novel shape map, SM_s: $CPF_s \rightarrow G_s$. This assignment is more explicitly expressed as follows:

$$S_{DC}(x, y, d, t) \rightarrow G(x, y, k). \tag{9}$$

Because two separate propagation layers are used for each shape (SL_1 and SL_2), there are also two different connection patterns. The specific correspondence between the 16 directions, d, and the specific synaptic connections, k, is made according to the two sets of directions as named above. The above process can be called a one-shot learning process and because it occurs with the centered shape, the connection pattern G therefore represents a centered CPF.

Recognition Recognition occurs by placing a learned shape onto the PM and letting it run through the orientation and DCs, whose activity is fed into each SM. The columns, however, serve then only to determine which (OC or DC) cell activity is fed into which shape layer, SL_1 or SL_2, of each SM: thus, the full range of directions is not exploited but merely used to classify them into the two sets of angles. Most SMs respond to most shape inputs to a small extent but the SM that bears the most similar centered CPF, will respond stronger than any other SM and thus signal the presence of the shape. To differentiate between the responses of each map, we monitor the entire spiking activity of each map during the time course of recognition. We call this the population activity, $P(t)$, of the SM and it is the sum of spiking activity of all neurons of both shape layers (SL_1 and SL_2) in that map:

$$P(t) = \sum_{x=1}^{W} \sum_{y=1}^{H} S_{SM}(x, y, t). \tag{10}$$

The parameters W and H are the width and height of the map.

Testing Five shapes were used, see Fig. 4: a rectangle, a circle, a triangle, a cross and a shape consisting of the superposition of the rectangle and the cross, which we now call square board. The first column is the centered shape, which was used for learning. The second column is a translated version of the shape, shifted to the lower right by 5 pixels in each dimension, which represents about 10 – 12% of the shape's width and height. The third column is scaled version of the (centered) shape, made smaller by 6 pixels in each dimension. The fourth column is the (centered) shape with a 'disturbance', a straight line extending from the right side of the image center to the upper right. The fifth column contains the shape made of a dotted contour, an input that is hereafter called fragmented.

The ability of a SM to identify was tested with its own centered, shifted, scaled (smaller), 'disturbed' and fragmented shape. The ability to discriminate was tested by placing the other, centered shapes on it.

3 Results

The operation of the system is demonstrated using the example of two straight orientations of different angle in a 15 × 15 pixel field (Fig. 5). The left two columns show how the SM for a vertical orientation responds to a vertical (preferred) and diagonal (non-preferred) input. The right two columns shows the reverse, how a diagonal-bar SM responds to diagonal (preferred) and vertical (non-preferred) input. The top row shows the CPF for the vertical and diagonal orientation (plotted twice each). There are no vectors in the proximity of any contours, because they were intentionally omitted: the initial period of the inward and outward propagating contours generates unnecessary vectors. The CPF patterns are then burned into the connection pattern of two separate, novel SMs (each having two shape layers) destined to be sensitive for only this propagation pattern. The bottom four rows in Fig. 5 visualizes how shape layer one (SL_1) responds to its preferred input. Time steps 2, 4, 5 and 6. Shape layer two (SL_2) does not contain any activity because no angles matched with those set of directions. At $t = 2$, the first contour input is dropped into the shape layer. Continuous input gradually increases the activity level through subthreshold propagation and eventually causes the shape layer to spike at $t = 5$, if the contour input matches the pattern connectivity (stored CPF). Spiking is marked as black pixels and those units 'turn white' in the next step due to the I&F reset. The diagonal input does not cause any spiking in the shape layer with vertical connectivity pattern (2nd column). The second and third column illustrate that a wave of 45 degree (or 135 deg) orientation is propagated exactly to its next, aliased orientation, skipping the one in between. The fourth column illustrates a limitation of the current architecture: there is spiking although the non-preferred (vertical) input has been fed to the shape layer. It can be seen at $t = 4$, that the activity is integrated across the diagonal direction and leads to spiking of every second unit. This does not happen for the 'reversed' case (2nd column) because a diagonal wave skips orientations (see 2nd and 3rd column again) and does not lead to integration in a horizontal connectivity pattern. Thus, there is a certain degree of overlap between some orientation pairs.

Moving on to the testing shapes, Fig. 6 visualizes the centered CPF_s for the rectangle and the circle. A CPF_s is determined from the first ca. 22 time steps. For shapes with closed contours, there is an inward- and outward-pointing vector field, for the interior and exterior space respectively; for shapes without closed contours, there is only an outward-pointing vector field, for example, for the cross (not shown).

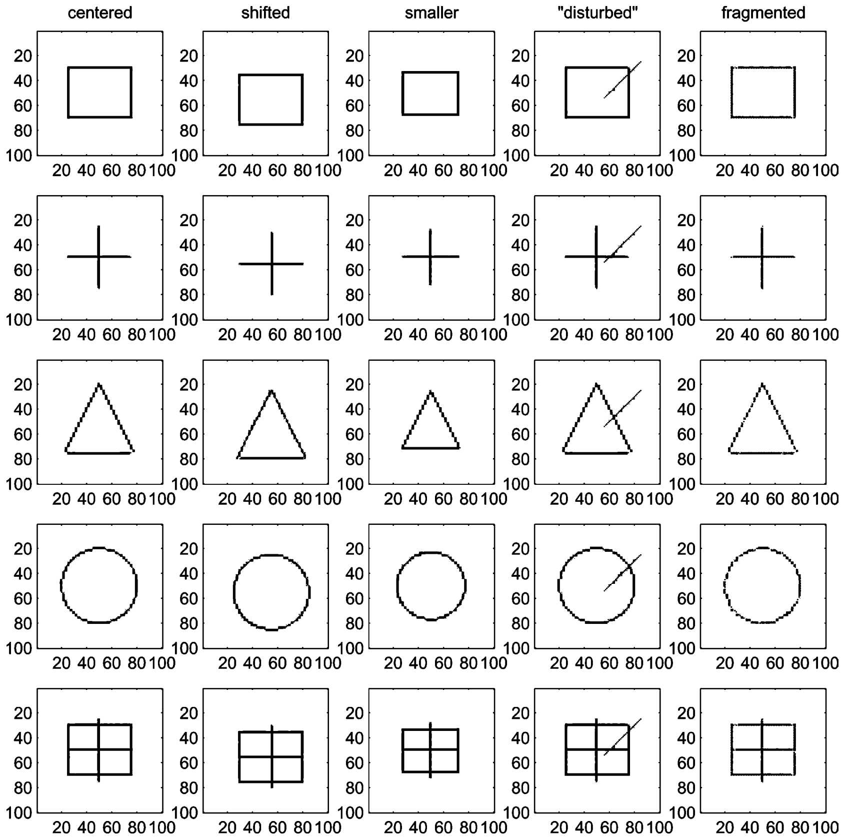

Fig. 4 Shapes used for testing. From top to bottom: rectangle, cross, triangle, circle, square board. The first column shows the centered shapes. The second column shows the translated shapes, which were shifted by 5 pixels into both axis directions (*shifted to the lower right*), which is about 10 – 12% of the shape width. The third column is a down-scaled (*smaller*) version of the (*centered*) shape. The fourth column shows the (centered) shape with a "disturbance", a straight line extending from the right of the image center to the upper right. The fifth column is a dotted version of the shape, called fragmented

As in the previous example, there exist no vectors in the proximity of the contours for technical simplicity. The vectors in the corner of the rectangle-CPF lie between the axes of the coordinate system, because the outward-propagating corner of the rectangle is rounded. Angles along the exact diagonals were sparse because at that location the corner is too sharp for any orientation to fit. The vector field for the circle is uneven at some spots: vector angles gradually change, but sometimes alternate back and forth. That is because the outward- and inward-propagating contours (on the PM) can locally change between angles due to the aliasing problem. This is especially true for vector angles close to either diagonal axis (45, 135, 225, 315 deg), which are only sparsely captured.

Figure 7 shows the response of the SM for the rectangle for its preferred, centered input (left column) and for a (centered) circle (right column). For the rectangle input, the horizontal and vertical (contour) sides cause the SM to fire (activity of SL_1 shown only). There is also integration along the diagonal but it is not sufficient to cause firing. The circle input causes the map also to fire, in particular those sectors with vertical or horizontal orientation. It does so with a delay for the inward-propagating contours, because they start further outside in the outward-pointing vector field for the rectangle, due to the slightly larger diameter (width) of the circle.

Figure 8 illustrates the summarized activity of the SM, $P(t)$ (Eq. 10), in response to various stimuli. The identification responses are discussed first.

The response to its 'centered' shape, called the signature now, is denoted as a thick, solid line: it rises steeply and stays well above the response activity for any other identification response. Its amplitude is determined by the length of the shape contour. It may increase or decrease during the propagation process depending on whether contours cancel each other out (e.g. square board) or whether they are only growing (e.g. cross). Part of the fluctuations are due to the aliasing problem (coarse nature of our network).

The response of a shape map to its 'disturbed' shape is shown as a thin, dash-dotted line, which also rises steeply but then runs below the signature: The 'disturbing' line causes a subtraction because it 'steals' a piece of area.

The response to the fragmented shape starts with a delay (of one time step) because it takes a short while for the PM to fuse all the contour pieces to one continuous (inward- or outward-propagating) contour.

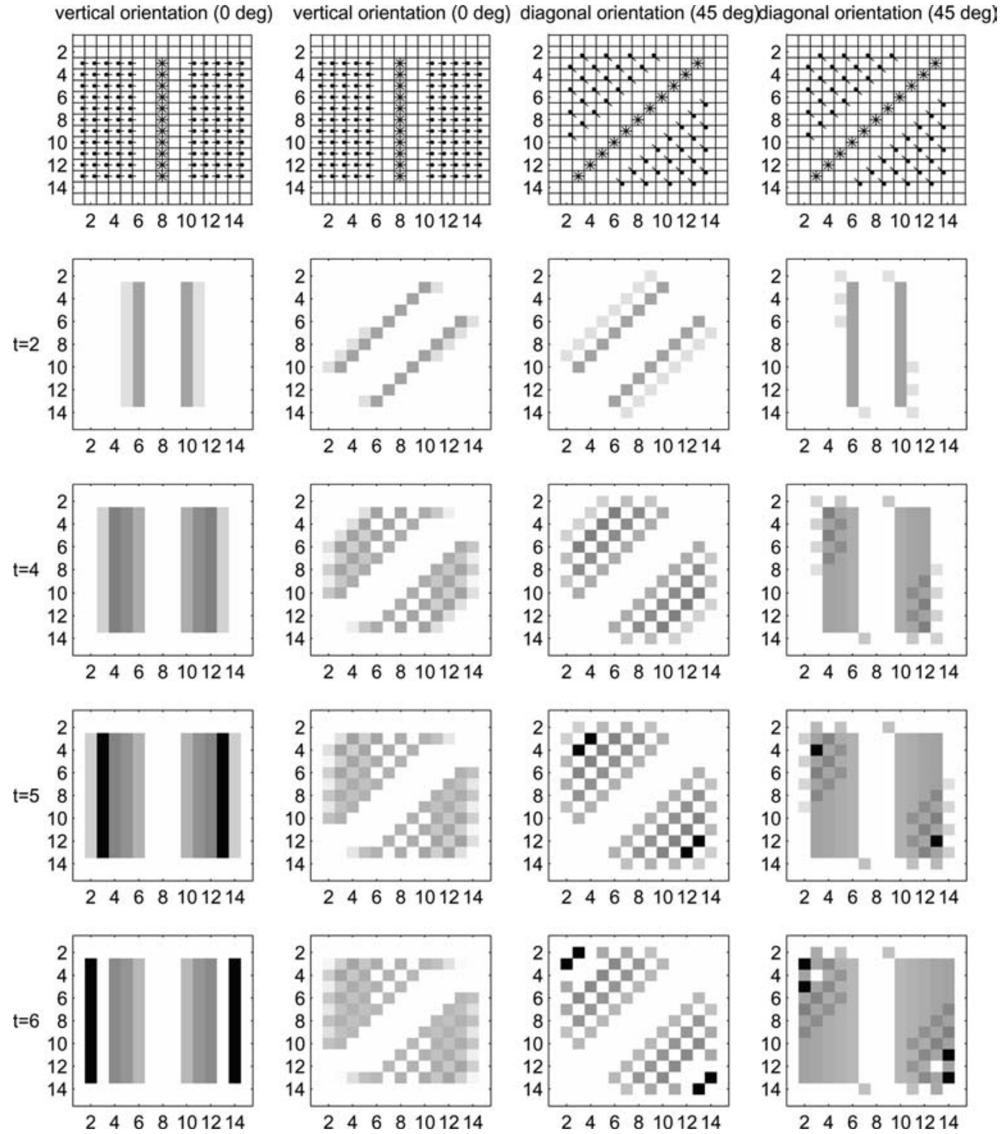

Fig. 5 Demonstration of the system on two orientations, vertical (0°) and diagonal (45°). Left two columns: the SM for the vertical bar was tested with its own preferred, vertical input (*leftmost column*) and a 45-deg bar (2nd column). Right two columns: the SM for a 45° bar was tested with its own, preferred column (3rd column) and a vertical bar (*rightmost column*). The *top row* shows the CPFs of the SMs (orientation indicated by stars): the direction of a vector is indicated by a point. The following *four rows* show the activity of shape layer one, V_{sm} of SL_1. *Gray-scale*: subthreshold values; *black color*: spike

The response to the down-scaled (smaller) shape starts immediately but reaches its peak stepwise because it takes a short while until both inward- and outward-pointing vector fields are covered, except for the cross shape.

The response to the shifted shape – plotted as thick, dashed line – is slower and reduced compared to the other identification responses. It also reaches its peak stepwise for the same reason as for the smaller shape. A translation by 10 pixels was also tested but those responses were not significantly different anymore from the discrimination responses.

The discrimination responses – the response to other (centered) shapes – are plotted as thin, solid lines. All those responses are mostly below and sometimes minimal compared to the identification responses. If, however, two patterns are similar in their exterior space, as it is the case for the rectangle and the square board, then they may have overlapping identification and discrimination response.

4 Discussion

4.1 Conceptual aspects

The presented architecture is a system that describes shape by the region that it outlines, rather than by mere contour geometry itself. The latter is often implemented as a contour integration approach, as it is, for example, pursued in virtually most neural networks (Rolls and Deco 2002). A contour

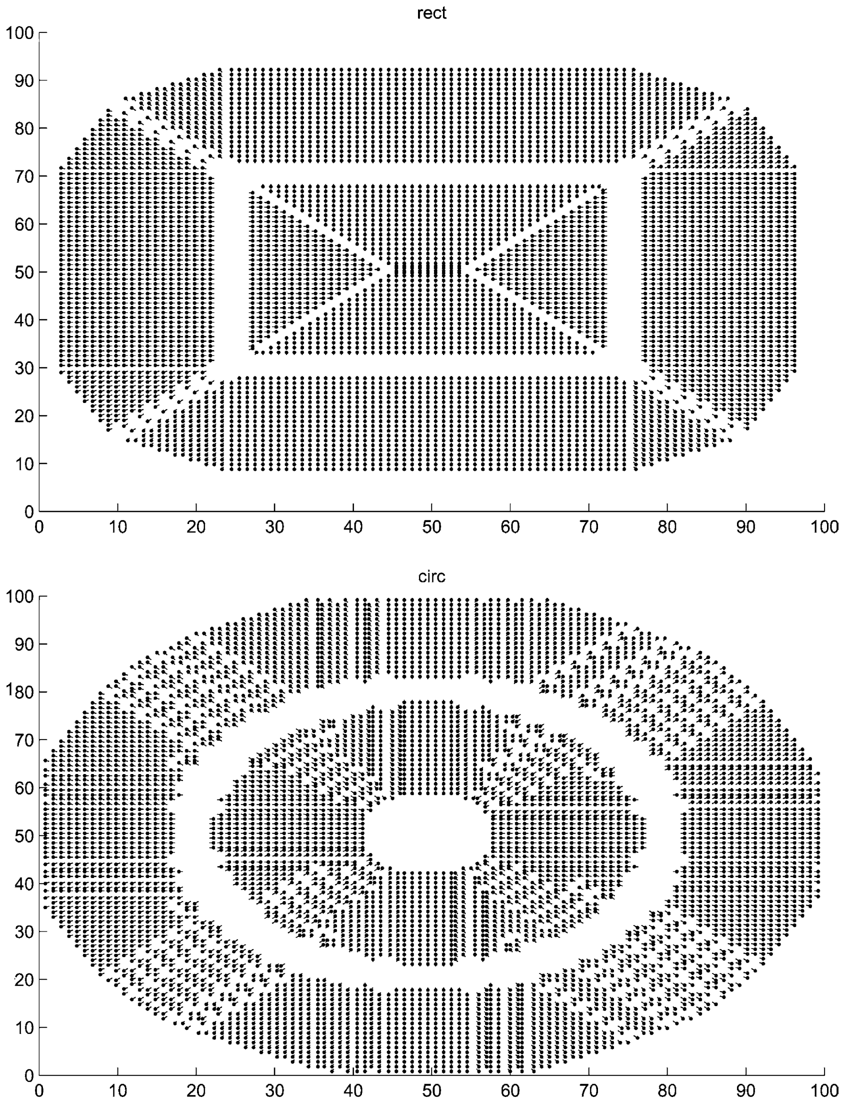

Fig. 6 Contour propagation field (CPF) of the centered *rectangle* (*top*) and *circle* (*bottom*). An inward-pointing and outward-pointing vector field is present

integration approach however suffers from the binding problem: at 'higher' stages of the recognition process, it is not clear anymore which contour belongs where. This binding problem is avoided by encoding the region. For example, the Fourier transform does encode region in some sense by determining the spatial frequencies. But small changes in a shape can introduce large changes in the power spectrum and the Fourier transform is therefore less suitable for encoding categories (or classes) (see (Palmer 1999 for a more elaborate discussion). A better solution might therefore be a contour propagation description. Blum's symmetric axis transform is one specific solution: the evolved sym-axes are highly useful for representation (e.g. Kovacs et al. 1998), but encode only the interior of a shape, which is only half of the space.

We have therefore proposed the above architecture, which senses the entire space inside and outside of a shape, thereby generating very distinct representations. The architecture can be labeled a template matching system, but the crucial difference to a traditional template matching system is that it senses space by making its input dynamic, a dynamic template matching system as one could call it. Sensing and encoding the space (or region) by contour propagation has the advantage of employing much more visual information than merely contour geometry itself. Judging from our first results presented here, we think that any shape can be easily distinguished from any other shape, as long as they are not structurally too similar. The architecture is therefore good for simple shape categorization: a learned shape can appear fragmented, scaled, disturbed, shifted to some extent and still be identified; likewise, the learned shape may have minor deformations or slightly different structure, it would still be recognized and hence categorized. Consequently, the architecture is less

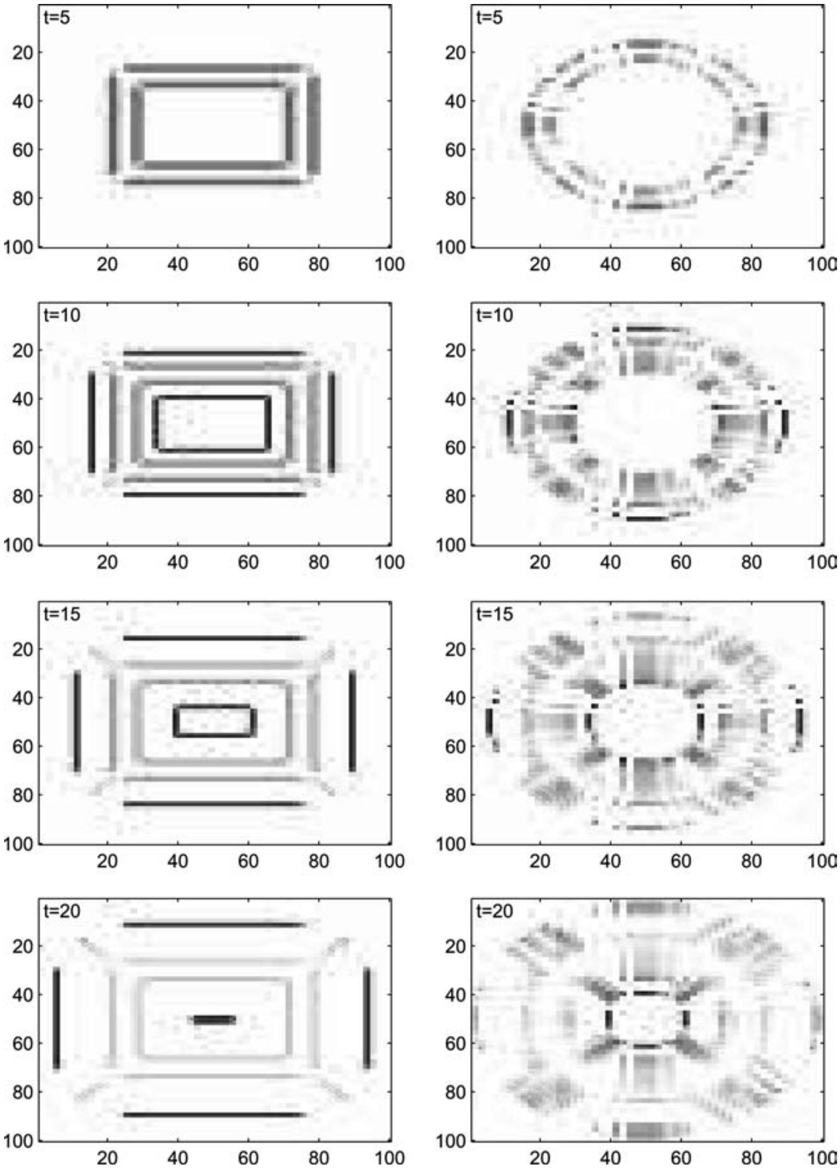

Fig. 7 Activity of the rectangle SM (specifically V_{sm} of SL_1), in response to its preferred, centered input (left column) and in response to a circle shape (*right column*). Snap shots taken at $t = 5$, 10, 15 and 20

suitable for an exact identification, for example, it could not discriminate between subtle shape differences. For that a differentiation mechanism has to be built in, that specifically learns minor structural details.

The architecture shows only a limited translation and size invariance. Although some studies show that the object recognition process is translation and size invariant to a large degree (e.g. Biederman and Cooper 1991), there are other studies indicating that this invariance may be more limited (see Dill and Fahle 1998 for a review). The difference in such psychophysical experiments may also result from different, specific conditions used. It may well be that (covert) attentional shifts contribute substantially to translation invariance in these experiments. The presented architecture does not capture any such shifts and is only able to recognize shapes that lie within the CPF, within the approximate focus in some sense.

4.2 Comparison and possible refinement

The present region-based description compares to a typical neural network (NN) performing shape recognition as follows (e.g. Fukushima and Miyake 1982; Riesenhuber and Poggio 1999; Amit and Mascaro 2003):

(1) The 'neural code' of a, NN is typically a rate code (Rolls and Deco 2002); recent networks employ timing codes (Hopfield 1995; Thorpe 1990). In our architecture, propagating waves are used, combined with single-spike detection of wave orientation and wave direction.

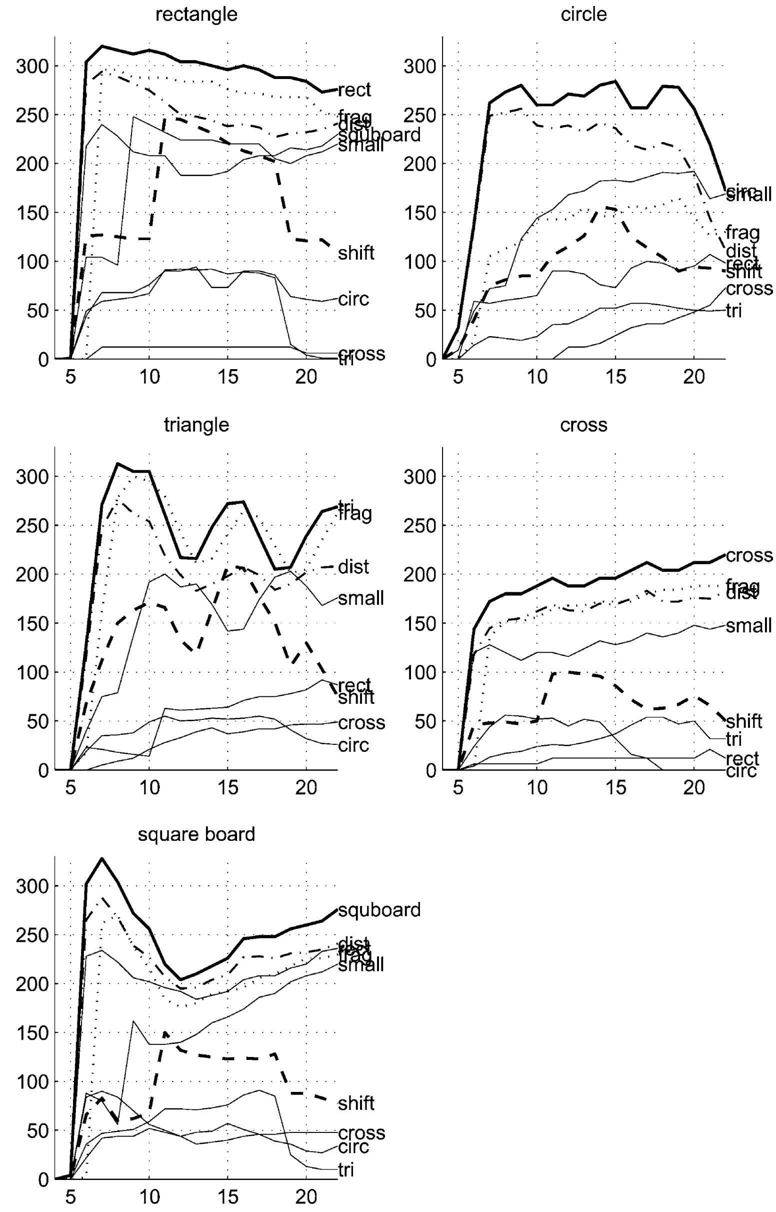

Fig. 8 Population activity, $P(t)$, during recognition evolvement for the five different shapes. X-axis: time. Y-axis: spike activity of the entire map (Eq. 10)

(2) An NN can hardly create such distinct (visual) shape representations, because the units in an NN tend to share information from different shapes, which leads to considerable overlap in representations.

(3) Learning in an NN requires a few learning cycles, whereas here it consists of a single cycle, namely the propagation of contours across the map.

(4) In an NN the recognition aspects of scale and size invariance are tackled by a local-to-global convergence and/or a fine-to-scale pyramid, which may lead to the loss of crucial information due to the merging of contour information or the coarsening of the resolution across hierarchical levels. Here, those aspects are partly solved by two processes: One is contour propagation, which allows for an almost perfect translation of contours over a limited distance. The other one is encoding space: even if the same CPF is compared to its translated, or scaled version, there is still considerable overlap between the two CPFs and thereof space, than between CPFs of different shapes.

Some of the NN are particularly tuned to carry out specific functions, like recognizing a few selected objects from different viewpoints (Riesenhuber and Poggio 1999) or the detection of objects by visual search (Amit and Mascaro 2003). The presented architecture is not designed for any such task although distorted shapes that represent rotated shapes in 3D space could be recognized to a certain degree due to the tolerance for structural variability. The architecture is especially good at learning any novel shape and recognizing it without fault despite some variations in position, structure, size and despite the presence of noise.

The present emulation is somewhat crude but, there are several places where one can improve.

(1) The actual direction-selectivity step is solved with a mere 'computer correlation' and had to be implemented with a neuronal mechanism to render this step biophysical more plausible.
(2) Although there is some overlap between certain angle pairs (Fig. 5, 4th column), this has not affected discrimination. Still, to refine the general ability to discriminate, it is desirable to distinguish between more orientations (and directions). To increase the number of orientations one could increase the resolution of the PM and probe it with larger RFs of, for example, 5 × 5 pixel RFs. To increase the number of distinguishable angles in the SM, one may also use a higher number of layers with different orientations, which – on the downside – may be too elaborate. Alternatively, one may make full use of the orientation columns: During the recognition process, they are only used to determine, which orientation activity is fed into which shape layer (SL_1 or SL_2). But it may be possible to use a single shape layer with a different connectivity that allowed for a larger number of orientations and a selective inhibition mechanism driven by the columns that allows propagation into one direction only.

The system is potentially capable of dealing with gray-scale images. Contours extracted from a gray-scale image are fragmented and a recognition system must therefore be able to deal with that. The example with the fragmented shape showed, that the architecture, in particular the PM has the ability to swallow fragmented input. Indeed, a next step would be to test the architecture on gray-scale images in connection with the retina developed previously (Rasche 2004).

4.3 Biological interpretation

We now discuss the biological plausibility of our architecture. The first question that may be raised is whether a region-based description exists in the brain. There is some psychophysical support that a region process like Blum's symmetric-axis transform takes place (Psotka 1978; Kovacs and Julesz 1994). Some neuroscientists have interpreted some of their neurophysiological recordings as an indication of some sort of region processing (Lee et al. 1998). Given the representational efficiency of the here proposed network one can take this study as a computational argument for the existence of region-based processes. The next question is then how such region processes may be implemented in the real visual system. One may assume that region processes are hard-wired: for instance Burbeck and Pizer presented a region-encoding mechanism in which units in lower levels would connect to units in higher levels to sense space (Burbeck and Pizer 1995). But one may take this a step further and suggest that contours are actually propagated in the nervous system. This may take place in the form of traveling waves, which have been observed in salamander retina (Jacobs and Werblin 1998) and in primary visual cortex of monkeys and turtles (Grinvald et al. 1994; Prechtl 1994; Prechtl et al. 1997, 2000; Bringuier et al. 1997; Senseman 1999). Most of these described waves are slow (millimeters/seconds) and may suffice, for example, to explain learning processes like the establishment of the CPF (although the turtle cortex does not have orientation columns). But for recognition, traveling waves had to be faster because recognition of shapes can occur blazingly fast (e.g. Thorpe et al. 1996). Fast waves have not been observed yet, but may also be difficult to discover (Glaser and Barch 1999).

The use of a vector field for representation was inspired by the existence of orientation-selective cells in the primary visual cortex (V1) covering the entire image. The majority of V1 cells are motion-sensitive and react to an oriented bar moving into a certain direction (e.g. Hubel and Wiesel 1968). In the current architecture, OC and DC cells are used as wave detectors. The OC cells sense the orientation of lines (or propagating wave fronts), whereby the full range of angles (0.360°) was used. As mentioned in the section, the term direction columns was invented and does not refer to any anatomical findings. No connections between columns were simulated.

Therefore, our biophysical inspiration is certainly somewhat loose, but our primary motivation was to find an efficient shape representation. What we specifically suggest is that one role of the orientation– and direction–selective cells may be to establish a CPF-like representation through wave propagation. The real (biological) V1 is a much more intricate network as compared to the one simulated here. It likely carries out a number of different roles by performing several independent perceptual analyses (Bruce et al. 2003). A biophysically realistic simulation of this architecture would have to account for several experimental findings, as for example, the pinwheel-like organization of orientations (Blasdel 1992a,b), their non-uniform distribution across cortex, the relatively constant distance between columns. The latter may even have an influence on translation and size invariance. There also exist substantial connections between cortical layers: none of these have been simulated here.

The here simulated SMs may exist in any visual cortical area. They do not necessarily have to be retinotopic but could be in a format that expresses different methods of shape encoding. Some neurophysiological recordings have shown that 'feature' columns exist in inferior temporal (IT) cortex, with neurons that signal for different shapes and forms (Tanaka 1996). Their activity could reflect the population activity of SMs.

4.4 Addendum

1. I realized later that the recognition process described in this study can be roughly thought of as a matching of

orientations. The CPF describes a three-dimensional matrix (the CPF matrix) with two dimensions corresponding to the spatial dimensions (x and y) and the third dimension being orientation. During recognition, the CPF matrix is cross-correlated against the orientation taken from the input (or testing) shape. Whereas this would describe an algorithmic way of thinking about this recognition process, the here presented implementation can be regarded as a biophysical plausible instantiation.

2. The idea to encode (2D) space has not been exploited to the extent as originally intended. That is because discrimination of simple shapes is not as challenging as the discrimination of basic-level objects or textures, which are generally more complex in their structure than mere shapes. To discriminate, for example, more complex shapes, we anticipate to encode more or even the entire space in order to make full use of the visual space (Rasche 2005a).
3. The study describes the simulation presented in Chap. 10 of (Rasche 2005a).

Acknowledgements The study was carried out in Prof. Michael Wenger lab at the University of Notre Dame funded by a grant from the University Graduate School. The author wishes to thank Michael Wenger for generous support.

References

Amit Y, Mascaro, M (2003) An integrated network for invariant visual detection and recognition. Vision Res 43(19):2073–2088

Biederman I, Cooper EE (1991) Priming contour-deleted images: evidence for intermediate representations in visual object recognition. Cognit Psychol 23(3):393–419

Blasdel G (1992a) Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. J Neurosci 12(8):3115–3138

Blasdel G (1992b) Orientation selectivity, preference, and continuity in monkey striate cortex. J Neurosci 12(8):3139–3161

Blum H (1973) Biological shape and visual science. J Theor Biol 38(2):205–287

Bringuier V, Fregnac Y, Baranyi A, Debanne D, Shulz DE (1997) Synaptic origin and stimulus dependency of neuronal oscillatory activity in the primary visual cortex of the cat. J Physiol (Lond) 500 (Pt 3):751–774

Bruce V, Green P, Georgeson M, Denbury J (2003) Visual perception: physiology, psychology and ecology. Psychology Pr, East Sussex

Burbeck C, Pizer S (1995) Object representation by cores – identifying and representing primitive spatial regions. Vision Res 35(13):1917–1930

De Valois RL, De Valois KK (1988) Spatial vision. Oxford University Press, New York

Dill M, Fahle M (1998) Limited translation invariance of human visual pattern recognition. Percept Psychophys 60(1):65–81

Fukushima K, Miyake S (1982) Neocognitron – a new algorithm for pattern-recognition tolerant of deformations and shifts in position. Pattern Recognit 15(6):455–469

Glaser D, Barch D (1999) Motion detection and characterization by an excitable membrane: The 'bow wave' model. Neurocomputing 26–7(jun):137–146

Grinvald A, Lieke E, Frostig R, Hildesheim R (1994) Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of macaque monkey primary visual-cortex. J Neurosci 14(5):2545–2568

Hopfield J (1995) Pattern recognition computation using action potential timing for stimulus representation. Nature 376:33–36

Hubel D (1995) Eye, brain, and vision. W.H. Freeman and Company, New York

Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol (London) 195:215–243

Jacobs A, Werblin F (1998) Spatiotemporal patterns at the retinal output. J Neurophysiol 80(1):447–451

Koch C (1999) Computational biophysics of neurons. MIT, Cambridge

Koffka K (1935) Principles of gestalt psychology. Harcourt, Brace, New York

Kovacs I, Feher A, Julesz B (1998) Medial-point description of shape: a representation for action coding and its psychophysical correlates. Vision Res 38(15–16):2323–2333

Kovacs I, Julesz B (1994) Perceptual sensitivity maps within globally defined visual shapes. Nature 370(6491):644–6

Lee T, Mumford D, Romero R, Lamme V (1998) The role of the primary visual cortex in higher level vision. Vision Res 38(15–16):2429–2454

McCulloch W (1965) Embodiments of mind. MIT Press, Cambridge, Massachusetts

Palmer SE (1999) Vision science: photons to phenomenology. MIT Press, Cambridge

Prechtl J (1994) Visual-motion induces synchronous oscillations in turtle visual-cortex. Proc Natl Acad Sci USA 91(26):12467–12471

Prechtl J, Bullock T, Kleinfeld D (2000) Direct evidence for local oscillatory current sources and intracortical phase gradients in turtle visual cortex. Proc Natl Acad Sci USA 97(2):877–882

Prechtl J, Cohen L, Pesaran B, Mitra P, Kleinfeld D (1997) Visual stimuli induce waves of electrical activity in turtle cortex. Proc Natl Acad Sci USA 94(14):7621–7626

Psotka J (1978) Perceptual processes that may create stick figures and balance. J Exp Psychol Hum Percept Perform 4(1):101–111

Rasche C (2004) Signaling contours by neuromorphic wave propagation. Biol Cybern 90(4):272–279

Rasche C (2005a) The making of a neuromorphic visual system. Springer, Berlin Heidelberg New York

Rasche C (2005b) A neural architecture for the symmetric-axis transform. Neurocomputing 64:301–317

Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2(11):1019–1025

Rolls E, Deco G (2002) Computational neuroscience of vision. Oxford University Press, New York

Senseman D (1999) Spatiotemporal structure of depolarization spread in cortical pyramidal cell populations evoked by diffuse retinal light flashes. Visual Neurosci 16(1):65–79

Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109–139

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520–522

Thorpe SJ (1990) Spike arrival times: a highly efficient coding scheme for neural networks. In: Eckmiller R, Hartmann G, Hauske G, (eds) Parallel processing in neural systems and computers. Elsevier, Amsterdam, pp 91–94