A Framework for Simulating Human Visual Recognition Suitable for Neuromorphic Implementation

Christoph Rasche
Institute of Psychology
Justus-Liebig Universität, Germany
rasche15(at)gmail.com

Boris M. Velichkovsky
Institute for Cognitive Studies, RRC "Kurchatov Institute"
Dresden University of Technology, Germany
velich(at)applied-cognition.org

Abstract

An algorithmic framework is presented, that is suitable for image categorization and scene understanding. The framework is based on a decomposition of visual structure into elementary contour segments, which then are synthesized to form structural relations and groupings. The output of the framework can explain all (structural) pop-out phenomena as observed in human visual search studies. This methodology is evaluated by carrying out a categorization task tested on image collections. The performance matches the one of computer vision approaches, but in contrast to such approaches, our framework allows to understand individual parts or even single contours of the image. A neurobiologically plausible model of this framework is envisioned, that can be implemented using neuromorphic networks.

Introduction

Visual recognition is typically initiated by firstly categorizing its input, i.e. by assigning the seen image to a basic-level category such as car, bird, office or tree. This assignment happens fast (within 150ms) and retrieves a memory structure with which a guided visual search using saccadic eye movements can be carried out. The saccadic search allows to fully understand the image by placing the fovea on selected objects and regions and to either verify hypotheses - which were made with completion of the categorization process -, or by even learning the peculiarities if it is a novel image. If one intends to build a neuromorphic system which can emulate such an image-understanding process then the primary goal should be to firstly model the categorization process. Without an understanding of this categorization process it is difficult to envision what exactly is required to build a saccadic visual search process.

There exist two respectable systems that perform this categorization to some extent. The system by Oliva and Torralba works on 8 super-ordinate categories from urban and natural scenes such as mountains, forest, street, or highway (Oliva and Torralba 2001). Their system uses the Fourier Transform to preprocess the image, whose output is then reduced by a principal component analysis, whose output in turn is discriminated by traditional classifier methodology. The system developed by Perona et al. (Fergus, Perona et al. 2007) also uses traditional computational methodology. It works well on sub-ordinate categories (leopards, soccer ball, cameras) shown from highly similar viewpoints. The short-coming with either system is that once the input image is categorized, the preprocessed image can not be used to interpret components/parts of the object or scene, that is, it can not be used to emulate a human visual search process using saccades. In order to do that one had to determine the geometry of individual contours for instance. For that reason, we pursue a categorization system which is based on contours and regions (relations between contours).

Neural networks (NN) also attempt to describe structure by integrating local orientation segments (Riesenhuber and Poggio 1999; VanRullen and Thorpe 2002; Hansen and Neumann 2004). However, this integration essentially corresponds to template matching as the contour is not transformed: merely the relative spatial location of the local orientations is stored. In contrast, our approach is based on a thorough parameterization of geometric structure (Rasche 2009). It distinguishes itself from other systems in the following ways:

1) The parameterization is obtained from transformations that act specific to the individual image structures and the resulting parameters can potentially be used to buffer structural variability. Structural variability describes the fact that different shape instances are often slightly different in their exact spatial relations and contour geometries. This variability appears as subtle and that is the reason why it is tempting to think that the variability can be absorbed through low-pass filtering – along the fine/coarse axis, which is implicitly assumed in neural network approaches. But low-pass filtering includes a loss of information and can therefore not lead to an accurate shape description. Instead, it is indispensable to analyze the structure without modifying it, specifically by making systematic distance measurements with increasing window size. From those we derive geometric parameters such as the curvature, smoothness, edginess and symmetry of a contour – and many more parameters for the relation between contours. Those parameters are then used as dimensions of a multi-dimensional space in which variability appears as a subspace.

2) The system can analyze any type of input such as texture, shape, object and scene, due to the through parameterization. It is *not* tuned to a particular set of categories or images as in the other approaches.

3) Our framework can explain a large number of psychophysical phenomena: a) the output of the structural analysis can explain all structural pop-outs as observed in visual search experiments (e.g. Treisman & Gormican, 1988); b) it allows determining high-curvature points of a structure as a human would; c) the system's categorization performance is only marginally affected by image rotation, similar to the rotation

invariance as observed for humans. Our approach thus uses similar transformations as the human visual system, if not even the same ones.
4) Our system offers the potential to perform a visual search that allows understanding parts or even individual contours of images – again due to the thorough parameterization.
5) Even if the image can not be categorized, because it is either a novel category or a category instance seen from a novel perspective, then the output of the structural analysis can be used to transform the structure to find the closest category or to rotate it 'mentally' to find the corresponding category representation for canonical views.
In the following we will sketch the present standing of our modeling efforts.

Framework

The framework is based on a *decomposition* of visual structure into elementary contour segments and areas (Figure 1). The starting point is the contour image CF, which is obtained from an edge-detection algorithm. From the contour image the contours are extracted with the goal to partition and describe them (labeled 'Local/global space'). The contour image is also employed to carry out a region analysis by contour propagation (labeled 'Symmetric Axes'). Both types of analysis will lead to elementary contour segments and areas; those structural elements are then integrated to form more complex descriptions such as structural relations and groupings, in a process that we call *synthesis*. In both processes, the structure is described in terms of geometric parameters; in addition, simple statistics of the luminance image are extracted for each descriptor (contour, area and its integrated descriptors), which we call *appearances*.

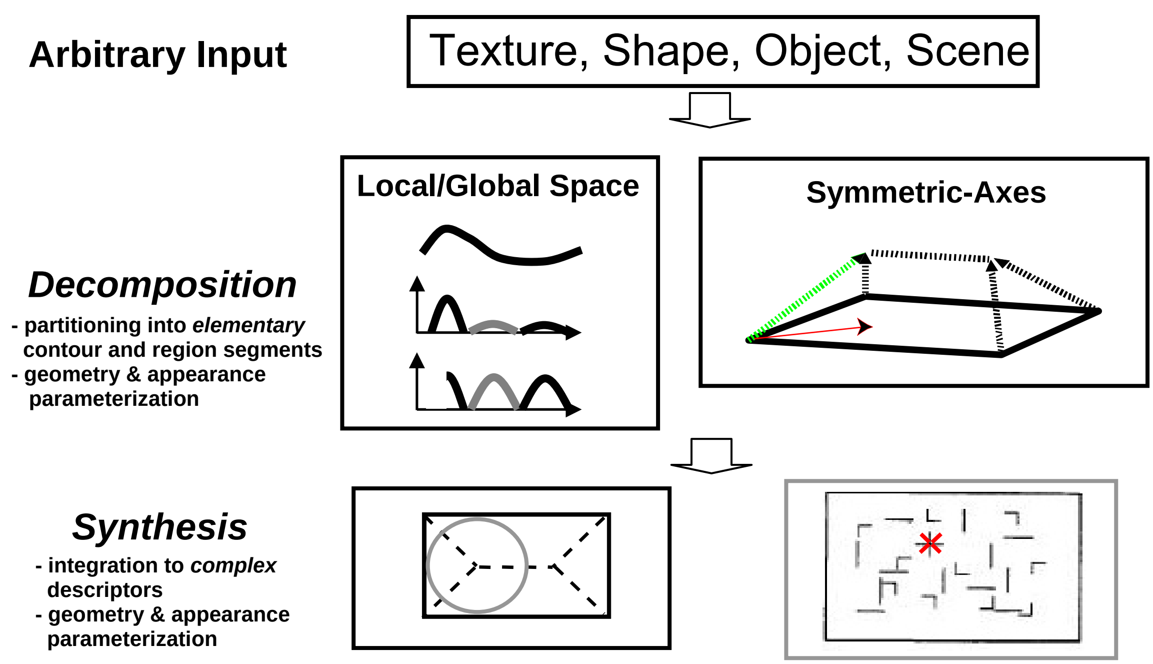

Figure 1. Key processes of the visual recognition process. **Arbitrary input**: The input image can be of any type. ***Decomposition***: A structure (contour image) is partitioned at points of highest-curvature resulting in elementary contour segments and regions. ***Synthesis***: The elementary contours and regions are integrated to complex descriptors such as intersections and groupings.

Decomposition

Local/global space.

The local/global space is an elaborate description of distance measurements made between points on a contour for different 'window sizes'. This is different from the 'scale space' as there is no filtering involved but just mere distance measurements and segment labeling. Specifically, a contour is iterated with a discrimination function which classifies a segment into arc (bow) or inflexion, leading to the 'bowness' and inflexion signature. The discrimination function acts on a window of fixed arc length (chord or stick) and determines the maximal distance between the contour segment and the straight line connecting the window's end points. If the selected segment lies on one side of the straight line, the maximal distance value will contribute to the bowness signature, but remain zero for the inflexion signature. The reverse applies if the segment lies on both sides of the straight line. For a range of window sizes the resulting signatures describe what we call the *local/global* space. In this space it is facile to localize high-curvature points by a maximum operation and it allows deriving parameters to determine the geometry of the partitioned segments.

This local/global space can be reduced to functions which describe the 'swinging' behavior of the contour. From these functions we can read out whether the contour describes an arc (parameter a), or whether is rather alternating (e.g. oscillating or wiggly; parameter x), and to what extent its curvature (b) is. In addition, one can derive the degree of edginess of the contour (parameter e) and the degree of symmetry (parameter s). To those 5 geometrical parameters, we add the parameters orientation (o) and length (l). In addition, we extract the following contour aspects, or *appearances*:

- Contrast (c_m, c_s): For each contour pixel, the range and standard deviation of its neighboring luminance values is determined. For a contour segment, these values are averaged resulting in values c_m and c_s respectively.
- Fuzziness (f_m, f_s): expresses the degree of fuzziness along the contour. Contours of natural scenes show often a high degree of fuzziness, in particular the ones lying within a textured region. Fuzziness is determined by preprocessing the gray-scale image with a DOG filter. Analogous to the contrast values, a contour mean and standard deviation is determined, but which are taken from a preprocessed image along the neighborhood of contour pixels.

- Isolation (r): quantifies the amount of region around a contour, or put differently the degree of isolation within a structure. Contours or shapes can appear isolated, for example the circle of a sun (or moon) in a landscape scene, or the rectangle of a picture in a room scene. This degree of isolation is characteristic as well and must be part of a contour description. It can be determined from the region analysis (see symmetric-axis transform next).

All these aspects are scalar values which are then combined to form the 12-dimensional vector $\boldsymbol{c}$:

$$\mathbf{c}(o, l, a, x, b, e, s, r, c_m, c_s, f_m, f_s).$$

So far we have been unspecific for what kind of contours the above description holds. The large variety of geometrical aspects promises a description of regular and irregular contours but a description of completely arbitrary contour geometries is not meaningful. In this study, contours are partitioned at points of U-turns (also known as 'ends'), specifically when the arc length of a segment for a given window size exceeds the length of a half circle with diameter ω (window size). The reasoning for this partitioning rule is that when contour segments lie opposite of each other as in case of an end, they are better described as two individual segments with a certain structural relation, which is exactly the idea of the symmetric axis (see next). Exemplifying this rule, an ellipse is partitioned into its 2 elongated arcs. The two segments of an L feature are also partitioned if they are of a minimum size.

Figure 2 shows the output of the contour decomposition in which straight and symmetric segments are marked as squares and diamonds respectively. Such segments can be easily determined by analyzing the local/global space and visually exemplify the success of contour decomposition.

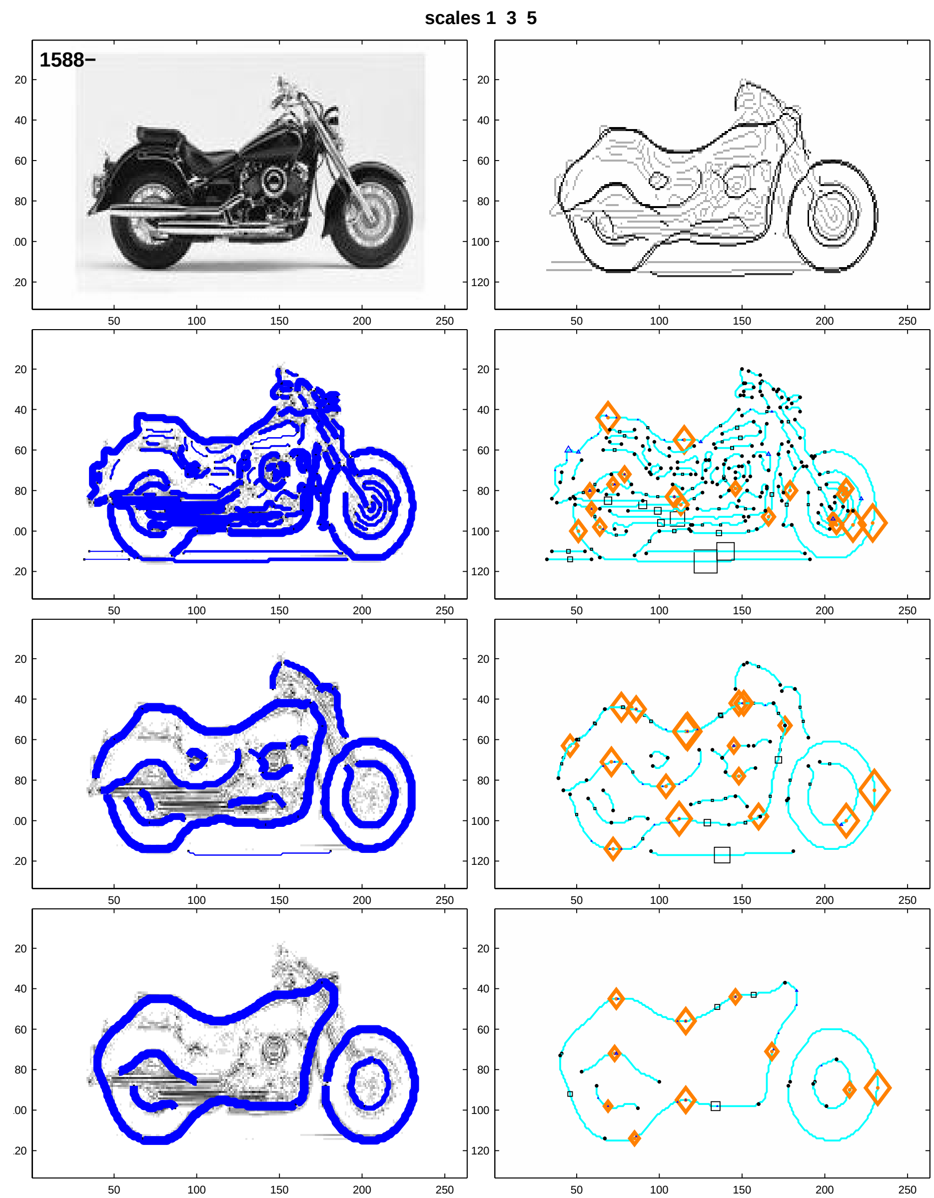

Figure 2. Example of contour decomposition (for scales σ=1, 3 and 5). **Left column**: Contour images with output of fuzziness filter (blob-filter). Contour thickness reflects mean contrast value; gray dots mark pixels with high fuzziness values. **Right column**: Geometric decomposition. Contour endpoints are marked as small black circles; squares and diamonds denote straight and symmetric segments respectively (their size reflects arc length – not curvature).

Symmetric-axis transform

To determine the structural relations between contours we use the symmetric-axis transform (SAT) by Blum (Blum 1973). The SAT produces a symmetric-axis (sym-axis) for a region by a contour-propagation process, typically resulting in several segments corresponding to trajectories or lines in 3D space (time as the 3rd dimension). There exist different instantiations of the SAT. For instance, Feldman and Singh provide an implementation which however requires multiple iterations and functions only on closed shapes (Feldman and Singh 2006). But in our framework, the SAT is rather used as in Blum's original proposal by implementation of a grass-fire process. It can be determined in two steps: the 1^{st} step consists of the propagation of contours whereby the temporal evolvement is held in a 2D map, also called the propagation field (or distance map in computer science); the 2^{nd} step consists of the convolution of this map with a high-pass filter with negative peak at the location of contour pixels (see Figure 3 for an illustration).

The sym-axes of a structure provide an elaborate description of the contour relations and need to be parameterized in order to compare them effectively. Toward that goal, the sym-axes are firstly partitioned into its constituent, elementary segments at the intersecting points (marked as plus signs in Figure 3). For example the sym-axis of a square is partitioned into 4 segments (half diagonals) describing each corner, or a rectangle's sym-axis is partitioned into 5 sym-axis segments with the central segment expressing the area between the two longer, parallel contours. A sym-axes segment is then parameterized as follows: the orientation (o) and length (l) of a sym-axis segment is determined, as it lies in the image plane. The initial and end distance is determined, e_1 and e_2 respectively, which are taken from the propagation field. And the degree of flexing (inward or outward bound segments) is determined and expressed with parameters c and p, which correspond to the distance and location. In addition to those geometrical parameters, the distribution of the luminance values of enclosed area is also characterized. The same contrast and fuzziness values are determined as for the contour (c_m, c_s, f_m, f_s), but are taken from the enclosed area (and not along the contour segments). In summary, we have the following 10-dimensional vector for a sym-ax segment, also called area:

$$\boldsymbol{a}(o, l, e_1, e_2, c, p, c_m, c_s, f_m, f_s).$$

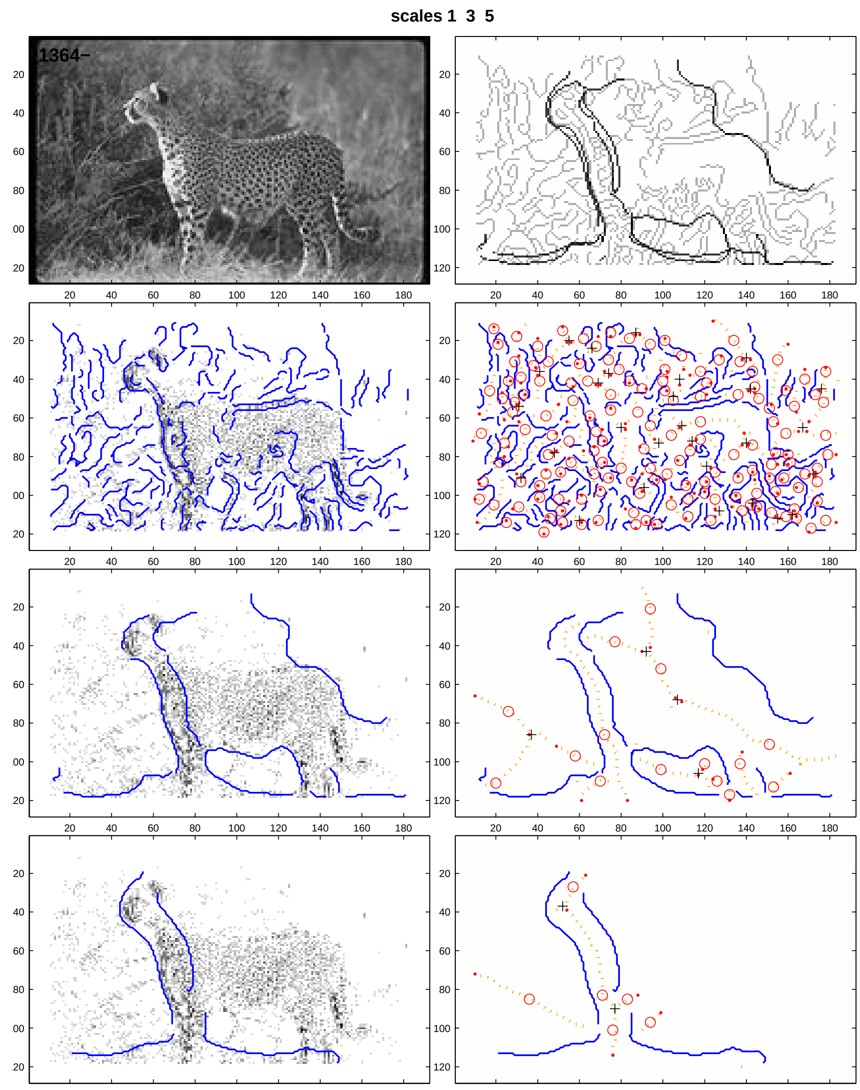

Figure 3. Example of region decomposition. **Left column**: as in figure 2. **Right column**: orange-dotted: sym-axes; plus sign: intersections of sym-axes segments.

Synthesis

The goal of the synthesis process is to form complex descriptions which represent potentially category-characteristic abstractions. As a typical contour image is very fragmented, such abstractions may be accidental. But we assume that the cognitive interpretation process can deal with this by usage of a 'top-down' process that extracts the essence of the image, namely its corresponding category – and not a detailed reconstruction.

To express more complex regions, as outlined by 3 or more contour segments, we also parameterized the region as spanned by intersecting sym-ax segments forming a 10-dimensional intersection vector:

$$\boldsymbol{k}(s_c, n, a, s_{min}, s_{max}, l_{min}, l_{max}, l_{mean}, l_{std}, \alpha_{min}, \alpha_{max}, c_m, c_s, c_r, f_m, f_s, f_r).$$

Dimension s_c is the symmetric distance value at the point of intersection; dimension n is the number of intersecting segments; dimension a is the spatial (2D) area of the entire structure; dimensions s_{min} and s_{max} are the minimal and maximal distance value for the distal (outer) ends of the intersecting segments; analogously, dimensions l_{min} and l_{max} are the minimal and maximal length of the segments; dimensions l_{std} and l_{mean} are the standard deviation and mean of the length values. In addition to those 11 geometric parameters, the same luminance and fuzziness parameters are added as for the area vector (c_m, c_s, c_r, f_m, f_s, f_r), taken from the area spanned by the intersecting segments, thus forming the above 17-dimensional vector.

Categorization Performance

To perform a categorization task, we firstly developed category representations by carrying out descriptor sorting: a descriptor (***c***, ***a*** or ***k***) was selected and all other descriptors of the entire collection ordered in decreasing similarity. If a selected descriptor would prefer descriptors in other images of the same category (for instance for the first 100 similar descriptors), then the selected descriptor was kept as a category-specific descriptor. To determine the corresponding for an image, its descriptors were matched against the list of category-specific descriptors and thresholded.

The performance of the system was evaluated on a variety of image collections.

a) Caltech collection (Fergus, Perona et al.): this collection consists of 101 subordinate categories in which images depict objects of relatively clear silhouette. We achieved a categorization performance of ca. 12%.

b) Corel collection: of the 600 image classes about 360 correspond to basic-level categories. We assigned them to 112 categories (for example, wild animals, patterns, sports, flowers, aircrafts, birds, cars, etc.). These images can consist of textures, objects or scenes. A categorization performance of ca. 11% was achieved.

c) Urban and nature collection by (Oliva and Torralba 2001). A categorization performance of ca. 40% was achieved.

It should be noted that these categorization results were achieved with completely unsupervised learning. The performance is comparable to computer vision approaches when they also use unsupervised learning. Though computer vision approaches typically report much higher categorization percentages, this is the result of supervised learning, e.g. a person pointing out salient object or scene characteristics to the system during the learning process.

The performance of our system was also tested when individual dimensions were omitted. If for instance, the orientation dimension was knock-out, the performance would drop by about half a percent. Systematic knock-out of individual dimensions showed that no dimension was substantially more important than any other one. The more dimensions were knocked out (as a group), the more performance decreased.

The framework could be refined and extended by a number of dimensions, which are useful for the basic-level and super-ordinate categorization process, as well as for the identification process.

1) The area outlined by the sym-ax and intersection vector has only been characterized by contrast and fuzziness values (c_m, c_s, f_m, f_s). But texture perception studies have shown that the detailed distribution of luminance values seems to be a strong determinant for proper texture identification (Dror, Willsky et al. 2004; Motoyoshi, Nishida et al. 2007). A parameterization of the luminance distribution may therefore be a better choice to describe the luminance values in an area.

2) The geometric contour parameters derived from the local/global space and spectrum are accurate but likely not precise enough for subordinate categorization or identification. For face identification for instance, the subtle contour geometry is decisive and must therefore be expressed distinctively, yet still show a certain degree of viewpoint independence. Such higher distinctness can be achieved by parameterizing the bowness distribution in the local/global space in greater detail, e.g. by determining their symmetry, skewness and flatness, analogously to determining the distribution for the luminance values.

Psychophysical and neurophysiological studies have shown that such high-precision parameterization can indeed take place (Riggs and Hunter 1973; Whitaker and McGraw 1998). Therefore, a more accurate parameterization of the sym-axes may lead to refined category representations.

Psychophysical Plausibility

The output of the framework can explain all (structural) pop-out phenomena as observed in human visual search studies (e.g. in (Treisman and Gormican 1988). The following figure numbers refer to Treisman and Gormican's publication and the feature display is given as pop-out vs distractor:

- Figure 2 (short versus long contour segments): the pop-out can be explained by a difference in the length dimension of the contour vector.
- Figure 3 bottom (pair of contour segments versus isolated contour segment): the pop-out can be detected by the presence of a sym-axis versus the absence of a sym-axis.

- Figure 5 (curved versus straight contour segments): the pop-out can be detected by the difference in the bendness dimension (*b*) of the contour vector.
- Figure 6 bottom (oblique versus vertical dimension): the pop-out is a difference in the orientation dimension of the contour vector.
- Figure 7 bottom row (ellipse versus circle): the pop-out can be explained by a difference in the length dimension of the sym-axes vector (l=0 for circles); there is also a difference in contour lengths, because a circle consists of a long, single contour, whereas an ellipse consists of two shorter contours (after partitioning).
- Figure 10 upper left (converging versus parallel): the pop-out can be explained by a difference in the sym-axis dimensions e_1 and e_2.
- Figure 10 center bottom (open versus closed L features): the pop-out is a difference in the dimensions *c* and *p* of the sym-axis vector.
- Figure 11 upper left (1/2 circle versus full circle): the pop-out can be detected by the presence versus the absence of a sym-axis.
- Figure 11 upper right (1/8 gap in circle versus full circle): the pop-out can be detected by the presence or absence of a sym-axis; or a difference in contour length.

The results of the categorization performance clearly show the representative power of the parameters when they are used as multi-dimensional vectors. The advantage of using a multi-dimensional space was proven with the robustness tests (dimension knock-outs). Despite the elimination of a few dimensions, categorization performance dropped only slightly, indicating that generally no single aspect is substantially more significant than any other one. This may explain why humans can recognize rotated pictures equally rapid as non-rotated pictures (Guyonneau, Kirchner et al. 2006): the presence of the remaining unaltered aspects may still allow for this rapid categorization. Nevertheless we hypothesize that eliminating dimensions will lead to prolonged categorization durations. Such a prolongation may be difficult to measure if only a single dimension is eliminated and if only a small data set is collected. We predict that with an increase in the number of eliminated dimensions, the categorization duration gradually prolongs, which could also be measured with data sets of regular size. The elimination of dimensions can be carried out by image modifications using computer vision methodology.

To accept this decomposition as a potential model of the early visual system, one probably has to firstly accept the requirement to make the variety of presented measurements on the structure, the generation of the local/global space for each contour, the derivation of a spectrum, the generation of a propagation field and the derivation of the symmetric axes and region image. But the multitude of parameters which are derived from those measurements, provide a simple explanation for the majority of the visual pop-out phenomena in Treisman and Gormican study (1988). Those pop-out effects are generally interpreted as supporting the traditional viewpoint, namely that the recognition process follows a gradual local-to-global integration along a hierarchy spanning several visual areas. But a newer viewpoint is that some form of global integration already takes place in early visual cortical areas using for instance horizontal connections amongst cells of the same neocortical layer, thus arguing rather for a global-to-local recognition evolvement (Kovacs 1996; Li 1998; Pettet, McKee et al. 1998; Hess and Field 1999; Rasche and Koch 2002). Some of those studies have suggested that SAT like processes occur, which we now have implemented and used to explain most pop-out phenomena. Given the fact that contours can lie arbitrarily in the image plane, we find neither type of hard-wired scheme adequate (Rasche and Koch 2002). We rather believe that flexible processes, such as the generation of the local/global space and the SAT, are better suited to transform image structure.

Neurobiological Plausibility and Implementation

What then could possibly be a neurally plausible computational substrate for the presented decomposition? The geometrical aspects of a contour could be read out from a histogram of the contour's local orientations: the distribution's width corresponds to the degree of curvature or circularity and the location of the maximum to contour's orientation. The use of such orientation histograms for structural description has already been suggested (Stevens 1978). Following this idea, a next step would be to include directional information to discriminate between arc and inflexion. The evidence on the existence of the SAT algorithms in the visual system has been sparse but persistent over the decades (Psotka 1978; Burbeck and Pizer 1994; Kovacs and Julesz 1994).

A neuromorphic implementation of this framework is envisioned, that can be emulated on a spiking neural network architecture such as the one worked out by Douglas' group, e.g. (Indiveri, Chicca et al. 2009). Parts of our architecture have already been simulated as neuromorphic networks, for instance the wave propagating process (Rasche 2007). But the networks that perform the parameterization still had to be developed. We imagine that the local/global space can also be emulated as a spiking neural network with the above suggested orientation histogramming. But one should not attempt to model this in a hierarchical fashion, as such an architecture does not provide the flexibility necessary to deal with the arbitrary layout of structure. The networks have to be dynamic and develop the parameterization independent of spatial location.

Discussion

Our quest of constructing a neuromorphic visual system was initially driven by merely developing the necessary algorithms that can solve the categorization process and by building the neural networks that can emulate the algorithms in neuromorphic networks (Rasche 2005). In this sense, we were following Marr and Poggio's viewpoint that there must exist an algorithmic solution to the computations performed in the brain (Marr 1982). With the discovery that our framework can explain structural pop-out phenomena, we begin to be particularly inspired by those visual search studies. They evidence, that a massively parallel structural analysis takes place, that can transform a structure independent of spatial location. It seems to us, that the idea of ‘self-collapsing shape’ - put forth by Gestaltists, e.g. (Koffka 1935) - is implemented in early visual cortical areas, but such networks have not been really pursued yet. As pointed out the key issue is to create networks which can perform the structural analysis translation independent. With the implementation of the symmetric-axis transform we have given a specific instantiation that should exemplify that direction.

The need for translation independence is also evidenced by Thorpe’s animal/non-animal discrimination task, in which two images are shown simultaneously in the parafovea (Kirchner and Thorpe 2006). An observer can make this discrimination within 120ms in average. One may argue that it does not require a complete categorization of both images to perform this discrimination; still, substantial structural processing needs to be carried out to arrive at a decision. One may also argue that attentional shifts are performed during that time, but this may be rather hectic. Instead, one should consider neural networks which transform a structure translation independent and which then signal the transformed structures by wave propagation.

As we can explain all pop-out phenomena, one may wonder whether we can perform fixation prediction for more complex scenes, e.g. in gray-scale scenes. This is a difficult task as the selected fixation locations by different observers can be rather variable due to the different motivations that the observers have. There exist approaches which attempt to point to those ‘hot spots’, such as the saliency map (Itti and Koch 2001). But these models can not recognize the structure at those points. However, there is much evidence that the human visual system selects its targets based on a structural analysis as demonstrated by studies on saccadic target selection (Richards and Kaufman 1969; Melcher and Kowler 1999). We therefore think that it is more important to understand the choice of preferred structures, than just a process pointing out potential spots.

There are several challenges associated with our approach, which need to be properly addressed in the future in order to build a perfectly functioning categorization and recognition system:

a) *Curse of dimensionality*: The decomposition and system produces a lot of parameters necessary to describe the enormous variety of categories. But many categories can be characterized by a fraction of these parameters and the abundant parameters may be rather detrimental to the performance. Thus, the large number of parameters needs to be better dealt with. A clever learning algorithm may solve the problem, but required a more systematic and incremental development of the category representation.

b) *Contour partitioning*: Proper contour partitioning is only possible with contextual information. So far we used only a partitioning of U-turns and L features. But that leaves still many contours potentially ‘category-unspecific’ and possibly not well discriminated. But to what extent this partitioning needs to take place needs to be tested systematically.
c) *Structural relations*: At the heart of an efficient categorization system needs to be a useful representation of structural relations (parallel segments, sequence of segments,…). Although this has been formulated many times, e.g. (Palmer 1999), there exists still no convincing method to represent and compare arbitrary structures. We think that the parameterization of structure and the formulation of a multi-dimensional space – as presented in this line of work – offers the best solution to it.

Acknowledgements

Our work is facilitated by grants from the European Commission (NEST-Pathfinder projects PERCEPT 043261 and MINET 043297, Network of Excellence COGAIN 511598), the European Social Foundation (Junior Research Group CogITo, Cognitive Interface Technology), the Russian Foundation for Basic Research (09-06-00293a and 09-06-12003obr_i) and the Russian Foundation for Humanities (09-06-01035a), to the second author. CR is supported by the EU Gaze-based Communication Project (IST-C-033816, under the Information Society Technologies Programme).

References

Blum, H. (1973). "Biological Shape And Visual Science .1." Journal Of Theoretical Biology **38**(2): 205--287.

Burbeck, C. A. and S. M. Pizer (1994). "OBJECT REPRESENTATION BY CORES." Invest. Ophthalmol. Vis. Sci. **35**(4): 1626-1626.

Dror, R., A. S. Willsky, et al. (2004). "Statistical characterization of real-world illumination." Journal Of Vision **4**(9): 821-837.

Feldman, J. and M. Singh (2006). "Bayesian estimation of the shape skeleton." PNAS **103**: 18014-18019.

Fergus, R., P. Perona, et al. (2007). "Weakly supervised scale-invariant learning of models for visual recognition." International Journal Of Computer Vision **71**(3): 273--303.

Guyonneau, R., H. Kirchner, et al. (2006). "Animals roll around the clock: The rotation invariance of ultrarapid visual processing." Journal Of Vision **6**(10): 1008--1017.

Hansen, T. and H. Neumann (2004). "Neural mechanisms for the robust representation of junctions." Neural Computation **16**(5): 1013--1037.

Hess, R. and D. Field (1999). "Integration of contours: new insights." Trends In Cognitive Sciences **3**(12): 480--486.

Indiveri, G., E. Chicca, et al. (2009). "Artificial Cognitive Systems: From VLSI Networks of Spiking Neurons to Neuromorphic Cognition." Cognitive Computing **1**: 119-127.

Itti, L. and C. Koch (2001). "Computational modelling of visual attention." Nature Reviews Neuroscience **2**(3): 194--203.

Kirchner, H. and S. J. Thorpe (2006). "Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited." Vision Research **46**(11): 1762--1776.

Koffka, K. (1935). Principles of Gestalt Psychology. New York, Harcourt, Brace.

Kovacs, I. (1996). "Gestalten of today: Early processing of visual contours and surfaces." Behavioural Brain Research **82**(1): 1--11.

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

Li, Z. P. (1998). "A neural model of contour integration in the primary visual cortex." Neural Computation **10**(4): 903--940.

Marr, D. (1982). Vision. New York, W. H. Freeman.

Melcher, D. and E. Kowler (1999). "Shapes, surfaces and saccades." Vision Research **39**(17): 2929--2946.

Motoyoshi, I., S. Nishida, et al. (2007). "Image statistics and the perception of surface qualities." Nature **447**(7141): 206-209.

Oliva, A. and A. Torralba (2001). "Modeling the shape of the scene: A holistic representation of the spatial envelope." Int. J. Comput. Vis. **42**(3): 145-175.

Palmer, S. E. (1999). Vision Science: Photons to Phenomenology. Cambridge, Massachusetts, MIT Press.

Pettet, M. W., S. P. McKee, et al. (1998). "Constraints on long range interactions mediating contour detection." Vision research. **38**(6): 865-79.

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

Rasche, C. (2005). The Making of a Neuromorphic Visual System. Berlin, Heidelberg, New York, Springer.

Rasche, C. (2007). "Neuromorphic Excitable Maps for Visual Processing." IEEE Transactions On Neural Networks **18**(2): 520-529.

Rasche, C. (2009). "An Approach to the Parameterization of Structure for Fast Categorization." International Journal of Computer Vision(accepted with minor revision).

Rasche, C. and C. Koch (2002). "Recognizing the gist of a visual scene: possible perceptual and neural mechanisms." NEUROCOMPUTING **44**: 979--984.

Richards, W. and L. Kaufman (1969). ""Center-of-gravity" tendencies for fixations and flow patterns." Perception and Psychophysics **5**: 81-84.

Riesenhuber, M. and T. Poggio (1999). "Hierarchical models of object recognition in cortex." Nature Neuroscience **2**(11): 1019--1025.

Riggs, L. A. and W. S. Hunter (1973). "Curvature as a feature of pattern vision." Science **181**: 1070-1072.

Stevens, K. A. (1978). "Computation of locally parallel structure." Biological Cybernetics **29**: 19-28.

Treisman, A. and S. Gormican (1988). "Feature analysis in early vision: evidence from search asymmetries." Psychol Rev **95**(1): 15-48.

VanRullen, R. and S. J. Thorpe (2002). "Surfing a spike wave down the ventral stream." Vision Research **42**(23): 2593--2615.

Whitaker, D. and P. V. McGraw (1998). "Geometric representation of the mechanisms underlying human curvature detection." Vision Research **38**: 3843-3848.