Chapter 29
Detecting, Representing and Attending to Visual Shape

Antonio J. Rodríguez-Sánchez, Gregory L. Dudek, and John K. Tsotsos

29.1 Introduction

In 1962, Harry Blum wrote a report titled "An Associative Machine For Dealing With the Visual Field And Some of its Biological Implications". The title reveals that he was not only inspired by, but also wished to impact biological vision. Blum was later motivated by the Gestalt psychologists in developing algorithms for extracting shape descriptors [4] and even tried to map his algorithm onto the results of Hubel and Wiesel's [21] study of visual cortical neurons. Blum points out that the Gestaltists used field theoretic concepts and proposed diffusion/propagation models. These ideas motivated Blum, but he realized they were unsatisfactory as presented due to their lack of precision and detail. Blum thus took those ideas and developed the now well-known Medial Axis Transform (MAT or 'grass fire' algorithm). The concept has reached its most sophisticated form in the shock graphs of Siddiqi et al. [48]. Our research looks at the detection and description of single object 2D silhouettes, the same kind of silhouettes on which MAT or shock graphs might operate. In our case, however, the quest is to develop a formalization of the stages of processing the primate visual cortex uses for this task and to show the correspondence between the computational result and the responses of single neurons to the same stimuli. In addition to constraining our design by the biological plausibility goal, we are further constrained by the quest to make the result amenable to attentional processes such as those required for spatial and shape reasoning [29, 56].

A.J. Rodríguez-Sánchez (✉)
Intelligent and Interactive Systems, University of Innsbruck, Innsbruck, Austria
e-mail: Antonio.Rodriguez-Sanchez@uibk.ac.at

G.L. Dudek
Centre for Intelligent Machines, McGill University, Montreal, Canada

J.K. Tsotsos
Dept. of Computer Science & Engineering, York University, Toronto, Canada

S. Dickinson, Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*,
Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5195-1_29,

Book ID: 302364_1_En, Date: 2013-05-21, Proof No: 1, UNCORRECTED PROOF

Shape computation in the primate visual system may be considered as part of the object recognition pathway covering areas V1, V2, V4 and the inferotemporal cortex (or IT) in the visual cortex. The first studies in area V1 found neurons that respond to bars and edges [22]. Already in those studies, three cell-types were differentiated: simple cells, responding to bars at specific locations; complex cells, which respond to a bar irrespective of its position inside the cell's receptive field; and hypercomplex (today known as end-stopped) cells, sensitive to the termination of an edge or a bar. End-stopped cells were extensively studied in later studies [2, 27, 33, 34], which reported the existence of end-zone inhibitory areas.

V2 neurons respond to real and illusory contours [57] as well as angles, corners, and provide submaximal responses to bars [6, 24]. V4 is important for the perception of form and pattern/shape discrimination [32]. The series of studies by Pasupathy and Connor [35–37] showed that populations of V4 neurons would respond to shapes and their responses could be approximated with an angular position-curvature representation of the shape. Posterior inferotemporal (PIT) neurons integrate contour elements with both linear and nonlinear mechanisms [**BriCon2003??**]. That study showed that some contours had an excitatory effect on the neuron response, while for others, it had an inhibitory effect. Anterior inferotemporal (AIT) neurons are responsible for the representation of objects, including faces, hands and other body parts. This representation includes shape as one of its components, this area receives inputs from V4 and PIT neurons at different retinal positions [52], which may explain its scale, position and view invariant cell responses [5]. <ref:??>

The developmental importance of shape is unquestionable [9, 17, 26, 44, 50, 51]. Spelke showed how in both adults and children, shape is an important component of object perception, and that Gestalt properties of shape are adhered to from a very young age. Smith et al. examined object name learning in young children (3 yrs) and found that learning object names tunes children's attention to the properties relevant for naming, namely, to the property of shape. Gershfokk-Stowe & Smith further showed this to be true for noun-learning in even younger children (17 months).

Finally, experimental work has clearly shown that humans and non-human primates can attend to shape [8, 10, 25, 45, 49, 54], and that this capacity interacts with other visual qualities or sensory modalities. Corbetta et al., using PET scanning, observed, that attention to shape activated the collateral sulcus, fusiform and parahippocampal gyri, and temporal cortex along the superior temporal sulcus. They concluded that selective attention to different features modulates activity in distinct regions of extrastriate cortex specialized for the selected feature. The disjoint pattern of activations suggests that perceptual judgments involve different neural systems, depending on attentional strategies. Todd, in a very nice survey paper, concludes that the perceptual representation of 3D shape may be primarily based on qualitative aspects of 3D structure that involve arrangements of salient image features, such as occlusion contours or edges of high curvature, whose topological structures remain relatively stable over viewing directions. He also points to empirical studies that have shown that the neural processing of 3D shape is broadly distributed throughout the ventral and dorsal visual pathways, suggesting that processes in both pathways are fundamental to human perception and cognition. Sereno & Amador found that during the presentation of a sample stimulus and test array to monkeys, some LIP neurons show stronger responses to the stimulus in the shape-matching task when the animal must attend to the shape of a stimulus, the first evidence that attention to shape can be seen in primate cortex. Cant & Goodale, using fMRI, showed that attending to shape activated the contour-sensitive lateral occipital (LO) area, whose organization seems complex, with neurons tuned not only to the outline shape of objects, but also to their surface curvature independent of contour. James et al. also found evidence that lateral occipitotemporal cortex (LO) is involved in representing object shape information. A specialization of LO, the tactive-visual area (LOtv) seems to integrate visual with haptic shape elements and even with auditory shape elements [25].

Although research on the detection and representation of shape has been strong over the years (see the chapters in this volume, for example), few shape models seem to support attentional processes beyond the usual region-of-interest kind of methods. A notable exception is the MetriCat model of Hummel & Stankiewicz [23]. It suggests two roles for visual attention in shape recognition: attention for binding and attention for signal-to-noise control. MetriCat implements both as special cases of a single mechanism for controlling the synchrony relations among units representing separate object parts.

Our goal is to develop a shape detection and representation methodology that supports the attentional processes as described by the Selective Tuning (ST) model of attention [55]. The choice of this model is that it includes a very broad set of attentional mechanisms and has already received very strong experimental support for the many predictions it has made regarding human and non-human primate visual processing [20, 55].

It is not difficult to use ST to constrain the quest for a shape detection framework. The requirements are all found in Tsotsos [55] and include both representation as well as processing constraints:

1. Visual representations (or areas to draw the direct comparison to cortical anatomy) are organized into a Lattice of Pyramids (or P-Lattice), defined in [55].
2. Receptive fields of individual neurons are spatiotemporally localized.
3. Objects, and their shapes, are presented using a parts-based composition of less abstract elements represented hierarchically in the P-Lattice.
4. The basic process of recurrent branch-and-bound operating over the P-Lattice is required for attentional tuning.

These are sufficient requirements for a shape representation scheme to be 'attentive' and thus play a critical role in the definitions of components that follow.

The next sections will briefly overview an early and then a very recent exploration into appropriate shape detection and representation ideas.

29.2 An Early Use of Curvature: Curvature-Tuned Smoothing

The original work on curvature-tuned smoothing (CTS) attempted to address this by representing shape in terms of curvature data and to allow a family of alternative interpretations via a nonlinear scale space [13, 14]. Since curvature is a differential property that must be inferred over noisy data, its extraction requires smoothing or regularization which, in turn, implies a biasing prior over the estimates to be extracted. The basis of the CTS approach is to employ a richer prior distribution than what is normally used. When one reflects on the importance of a prior, it is only a small step to realize that top-down influence can be used to moderate or accelerate the estimation process, a step that was not taken in the original work on curvature-tuned smoothing which was based on exhaustive consideration of all possible curvatures, but which relates to later work on attentive processing.

The perceptual relevance of curvature, particularly for 2D curves, has been apparent for decades while the use of a multi-scale representation sidesteps the issues of more simplistic representations. In prior work, the stable extraction and measurement of curvature information in the presence of noise was addressed in several ways, but was usually based on the assumption that there is a single unique curvature measurable at each point. While this is, of course, true in the analytic case, the assumption introduces significant difficulty for estimation problems involving noisy signals, such as those that occur in vision. Despite the respectable results that have been achieved by some researchers, the need for scale-specific operators to deal with noise problems (which also manifests itself as the need to choose a best smoothing scale, or the choice of an appropriate neighborhood for measurements) causes an inherent preference for certain ranges of curvature value and involves strong implicit assumptions about the underlying signal. The actual curvature of a signal depends on what we call noise and what we call signal, and hence may take on differing values depending on our goals.

The extent and shape of the neighborhood used for this processing asserts an implicit scale specificity as a result of the interpolant of support function used for estimation. For example, a polynomial model of a portion of a curve limits the number of inflection points over the region and hence bounds the amount of structure that can occur. In general, high curvatures with correspondingly small spatial extents relative to the neighborhood size will be lost or drastically attenuated. This attenuation is, in fact, the key objective of the non-local estimation methods. On the other hand, low curvatures may remain difficult to measure since the neighborhood being used will often be too small to reduce local noise. To a large degree, this too is the objective of non-local modeling: to discard structure at the wrong scale. The difficulty is compounded in practice by the fact that scale-specific constraints are usually stated only implicitly and the single correct scale is difficult to control or select. In most modeling problems, the objective is to map the data to its most likely causative models, that is, the most reasonable real curves that it could actually describe. In doing so we regularize the measurement process, discarding implausible structure in the data. The method described here exploits the relationship between curvature and scale to produce a set of alternative descriptions of the data based on structure at different scales.

Our approach begins with shape primitives that are extracted using a variational formulation called *curvature-tuned smoothing* [12–14]. This description has several desirable properties including its basis in perceptually-relevant curvature measurements [1, 28], and its properties in the face of sparse data or noise [38, 53]. The multi-scale nature of the representation allows multiple alternative possible descriptions for portions of a curve to be retained. It produces a description of a curve where a single region may be described in terms of one or more arcs of different curvatures (and hence sizes), and hence makes explicit information and different spatial scales (by the term scale we refer to the size or spatial extent of a processing operation or feature).

The curve representation is produced by repeatedly minimizing the following energy functional with respect to a piecewise C^2 solution $\bar{u}(t) = (x(t), y(t))$:

$$E\big(\bar{u}(t,c)\big) = \int_{l_e}^{k_e} \big\|\bar{u}(t) - \bar{d}(t)\big\|^2 + \phi p\big(\bar{u}(t)\big) + \lambda(c)\big(\kappa_a(t) - c\big)^2\, dt$$

where t is arc length, $\bar{d}(t) = (x(t), y(t))$ is a list of initial data points estimating the input curve, $p(x, y)$ is a potential function derived from the input image (i.e., a measure of edginess), with ϕ being an associated weight, $\kappa_a(t)$ is the curvature of $\bar{u}(t)$, $\lambda(c)$ provides the relative weight of the stabilizing term, c is the "curvature tuning", and I is the stabilizing constant selected as a function of c. The term ϕ can be set to zero if pure 2D curves are the input data (as opposed to edges embedded in a larger image). This solution is determined for various values of c, denoted by c_i. The first two terms constrain the solution to be consistent with an initial input description and with image support for the curve position. The third term expresses an "internal" bias for a solution with a specific curvature given by c. The result is a multi-scale decomposition of a curve such that segments that can be interpreted as being characterized by different natural curvatures are simultaneously extracted. These are the regions having low energy in terms of the above functional. An example of the result is shown in Fig. 29.1. The figure shows a poison sumac leaf in silhouette and the portions of it that are detected at specific curvature tunings along the silhouette.

The matching methods most commonly used for curved data deal with recognition by organizing cues along the arc-length axis. That is, a correspondence between features is established as the curve is traversed in a given direction. The presence of structure along the curvature (non-linear scale) dimension is an additional and unique aspect of the description produced from curvature-tuned smoothing. For example, the leaves of the poison sumac plant are typified by large rounded leaf tops containing a particular arrangement of three "sub-bumps" at the same location.

By using the multi-scale representation to match curves in scale space, a potentially richer description was obtained that what would be extracted by comparable regularization-based smoothing techniques. These multi-scale descriptions could then be used for recognition, for example using dynamic programming [13]. Most notably, this representation using various prior expectations in curvature space can "tune" the regularization process. Whether this tuning should be applied selectively instead of exhaustively was never explored in the original work, but is a natural

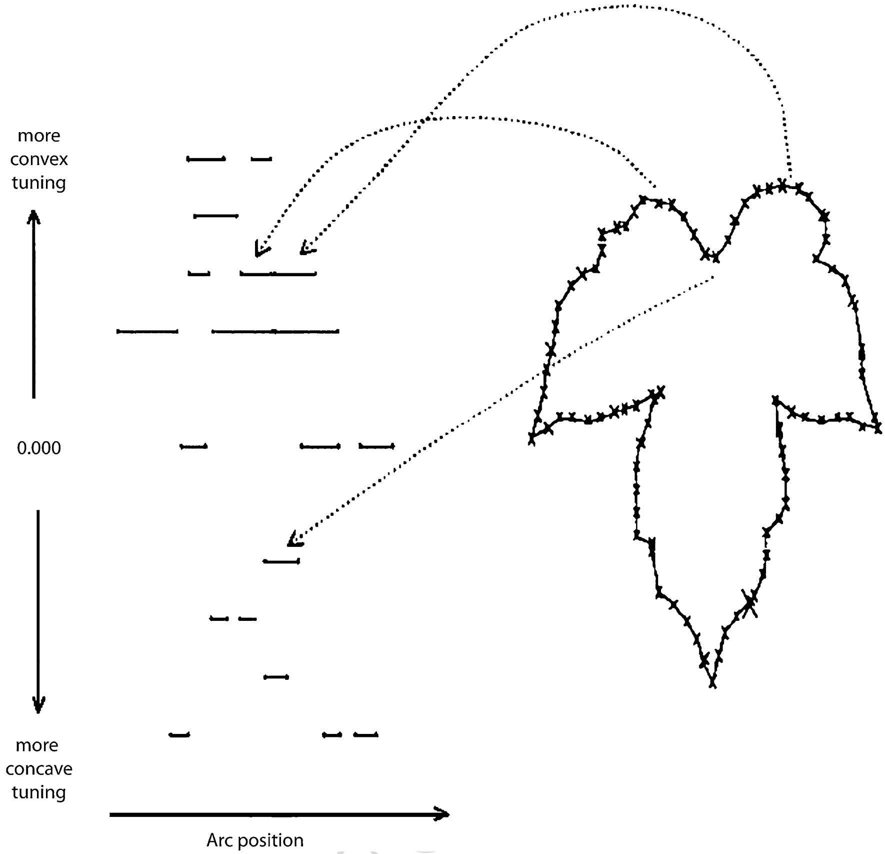

Fig. 29.1 Poison sumac leaf and scale-space. The CTS description of the poison sumac leaf is shown, with the segments corresponding to certain features on the leaf illustrated. Each line corresponds to a segment with discontinuities at its ends. The length of each line corresponds to the segment length

candidate of top-down bias in the interests of either computational efficiency of selective search and thus a natural hook into attentional processes.

29.3 2DSIL: End-Stopped and Curvature Computations for Silhouette Recognition

Our most recent efforts have focused on trying to create a shape model with biological relevance if not also plausibility. Recent experiments in area V4 [37] and TEO [7, 52] of the macaque monkey seem to agree with a recognition of objects by parts strategy, clearly suitably satisfying for constrains the ST attention model. In the case of V4 and TEO, those parts would be local curvatures [7, 35–37]. 2DSIL

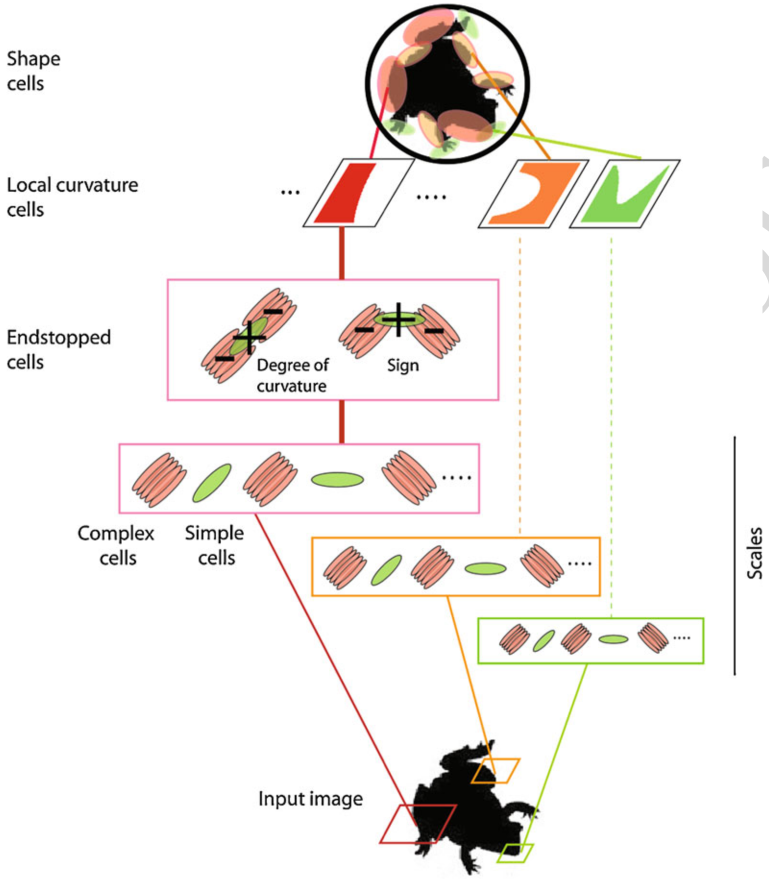

Fig. 29.2 Architecture of 2DSIL (see text and [42, 43], for more information)

[43] (see Fig. 29.1) is our resulting model. Different from other models, such as [39, 47], 2DSIL does not consist of the addition of new layers over the Neocognitron [16] with a repetition of S and C neurons. Rather, new types of neurons select for different curvatures and include inhibitory surround. Cell types comprising 2DSIL (Fig. 29.2) are the following:

- *Simple cells* of visual area V1 are sensitive to bar and edge orientations. Gabor filters [31] and Difference of Gaussians have been shown to provide a good fit when modeling simple cells from area V1, although a better fit to neuronal responses has been found with Difference of Gaussians [19]. The latter formulation is the one used in 2DSIL for modeling simple cells. 48 different groups of simple cells were designed, varying sizes, orientation and values of Gaussian width and length.
- *Complex cells* have a sensitivity for bars and orientations as well, but their receptive fields are larger than those of simple neurons. Hubel and Wiesel [21, 22]

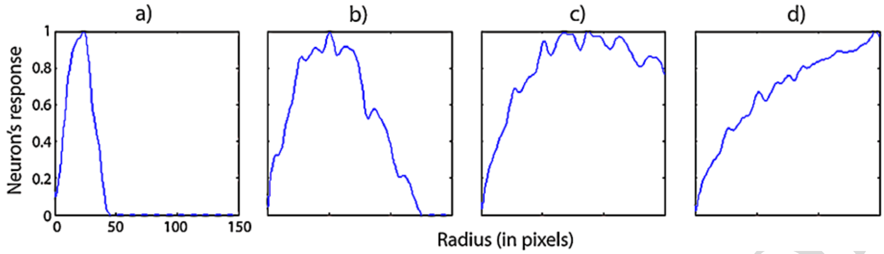

Fig. 29.3 Curvature selectivity from end-stopped neurons. Smaller cell sizes (**a**, **b**) are selective for sharper curvatures, larger neuron scales are selective for broader curvatures (**c**, **d**). Simple cell sizes that combined into end-stopped cells were: 40 (**a**), 80 (**b**), 100 (**c**) and 120 (**d**) pixels

found that simple cells have one or more subfields in which the response is either on or off while complex cells yield both on and off responses, which suggest that complex cells integrate the responses of simple cells. In our model, a complex cell is the sum of 5 laterally displaced model simple cells Gaussian weighted with position and later rectified (any value less than 0 is set to 0).

- *End-stopped cells* can be of two types. One provides band-pass selectivity for degree of curvature. The tuning for degree of curvature can range from very sharp to very broad as can be seen in Fig. 29.2 for four cell sizes. This type of cell is composed of a simple and two complex cells [11]. Complex cells are laterally displaced and provide an inhibitory input with respect to the centered excitatory simple cell. Depending on the orientation of the complex cell component with respect to the simple cell we obtain neurons that are selective to degrees of curvature (if that orientation is the same). The combination from smaller model end-stopped neurons is selective for sharper curvatures and the combination of larger cells responds strongly to broader curvatures (Fig. 29.3). The second type of end-stopped neuron is selective for the sign of curvature, by using displaced neurons at different orientations (Fig. 29.2).
- *Local curvature cells* are obtained due to the neural convergence of the two types of model end-stopped cells. By combining model end-stopped cells selective to the degree of curvature and model sign end-stopped cells responses, we obtain twice the number of curvature classes than the number of end-stopped cells. For example, if we have four types of degree of curvature end-stopped cells, through the use of the sign of curvature of those cells we obtain eight curvature classes. For the case where the response from end-stopped cells is small, a high response from a model orientation simple cell means the contour is a straight line, so its curvature is set to 0. Local curvature cells are computed at each location.
- *Shape cells* are at the top of the hierarchy (Fig. 29.2) and integrate the responses from local curvature cells. Shape-selective cells respond to curvature configurations with respect to their position in the cell's receptive field. A model shape cell will respond to a shape, and depending on how close the stimulus is to its selectivity, its response will be stronger or weaker. In the example provided in Fig. 29.2, the input to a shape cell that respond to the silhouette of a frog is composed of

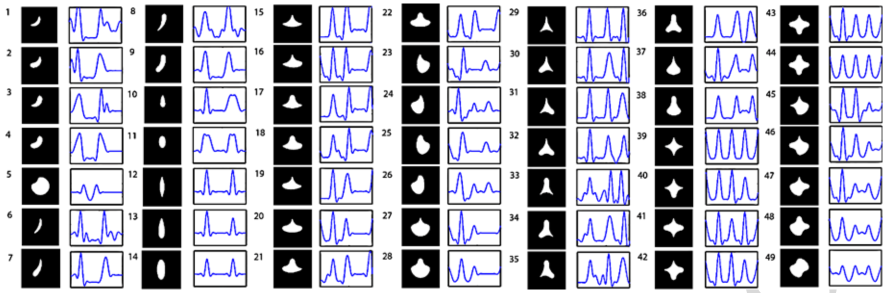

Fig. 29.4 Capability of shape neurons for encoding stimuli from Pasupathy and Connor [37]. Stimuli (in *black background*) were created using a Matlab program for that purpose provided by Dr. Pasupathy. Compare the plots at the right of the stimuli with the neural responses and plots in Fig. 3 from Pasupathy and Connor [37]

local curvature cells with high responses to sharp curvatures at the bottom (the right hand of the frog), local-curvature cells selective to broad curvatures at the left and top-left (two back legs), etc., providing a cell that has a high response to different local curvatures at specific locations. A similar shape would also provide a high response from the 2DSIL shape-selective cell.

2DSIL shape-cell responses were compared with the responses from neurons in area V4 [43]. Neurons in area V4 of the visual cortex encode shapes as curvature parts relative to their position in the object [37]. The stimuli used in that study were silhouettes created using convex and concave boundary elements to form closed shapes (see Fig. 29.4, silhouettes on black background).

Figure 29.4 shows the results of applying 2DSIL over the stimuli (left columns) from [37]. The encodings from model shape cells are in the right columns. The blue plots not only reproduce the curvatures for the stimuli that appear at their left but are also very close on how populations of V4 neurons encode shape, compare this figure with Fig. 3 of [37] or refer to [41]. When computing the difference from the plot values in Fig. 29.3 with those of [37], the reported error was of 0.074 ($stdev = 0.037$, $error\ range = [0, 1]$) which shows that the model shape cells in 2DSIL faithfully replicate the population results obtained in area V4 of the visual cortex.

We further tested 2DSIL on real images. We selected eight commonly used databases with clutter (Leaves from Fergus et al. [15], cars back, faces, motorcycles, leopards, bottles and airplanes from Caltech256, and cars from Leung [30]). The task was an object present/absent classification, where the model has to detect if the object in question is present in the image or not. We used the background database as negative (absent) samples.

The details of the test have been presented previously [42]. The key here is to simply show that the curvature cells in the model do indeed capture sufficient salient aspects of shape to enable classification. Values from local curvature cell responses were used to construct a feature vector (2640 elements) that was the input to an Adaboost classifier (300 iterations). Training consisted of presenting randomly half the images containing the object (positive samples: 93 for leaves, 263 for cars back, 258 for cars-MIT, 225 for faces, 95 for leopards, 50 for bottles, 413 for motorbikes and 537 for airplanes) and half the background images (negative samples: 225 randomly chosen images). The remaining images were used to test the model (same number as in training, but different randomly chosen images).

We obtained the percentage of correctly classified images (as containing objects or background). The model outperforms classical systems such as for most databases. Correct classifications were: 98.6 % for cars back (1.9 % false negatives and 0.9 % false positives), 96.9 % for cars-MIT (5 % false negatives and 0.9 % false positives), 89.2 % for faces (12 % false negatives and 10 % false positives), 94.0 % for leaves (10 % false negatives and 0.7 % false positives), 96.9 % for leopards (4 % false negatives and 2.6 % false positives), 83.3 % for bottles (35 % false negatives and 16.5 % false positives) and 92.8 % for airplanes (5 % false negatives and 1.2 % false positives). Results are similar as well to another biologically inspired model [46], and the very recent Bag-of-features approach by Han et al. [18].

Finally, since the ability to connect to an attentional system such as Selective Tuning provided key constraints for the overall design, it is important to show that these constraints are indeed satisfied. In Rodriguez-Sanchez et al. [40], we showed exactly this capacity demonstrating how the shape cells provide sufficient information for simple shape recognition in common visual search tasks. The performance of the overall shape attentive system was directly compared to psychophysical experimental data in common search tasks: a color similarity search where feature search can be inefficient if the differences in color are small and a set of feature and conjunction searches that show the continuum of search slopes from inefficient to efficient using stimuli such as circles, crosses, and letters. It was shown that the qualitative performance comparison was virtually identical.

29.4 Conclusions

Our foray into shape representation, detection and attentive recognition, has led to a sophisticated and successful model, 2DSIL, of processing in the early stages of visual cortex and also to a high performance computer vision shape framework. This work, however, suggests as many questions as it might answer. Questions that motivate the next stages of research include:

- How would higher order processes use 2DSIL as input, such as those examined by Brincat & Connor [7]?
- Can the model be extended to surfaces or 3D shapes, and precisely how? Although the CTS model was extended to operate over range data, how might it be applied to natural imagery with implicit 3D structure, and how could this extension be made for 2DSIL? Moreover, while curvature extrema regions of constant curvature and vertices are both computationally natural primitives with extensive evidence with respect to perceptual relevance, the choice of tractable yet perceptually-relevant descriptions for surfaces is much less clear. Despite extensive evidence for the importance of 3D structure, are the mathematically or com-

putationally elegant model extensions of 2D shape suitable for modeling human perception?

- Several researchers have reported selectivity for 3D shape in IT [**JanVogetal2000??JanVogetal2001??DurNeletal2007??VerVogetal2010??**]). The lower <ref:??> bank of STS (superior temporal sulcus—a subarea of TE) was found selective to 3D shape, while lateral TE was selective to 2D shape [**JanVogetal2000??**]. How <ref:??> in the context of 2DSIL, can local curvature neurons be extended from curves in 2D-silhouettes to surfaces and shape cells to encode from shapes in a plane to shapes in 3D space [**OrbJanetal2006??YamCaretal2008??**]? <ref:??>
- How can the model, which permits all potential shapes, be tailored via learning to represent the set of real objects in a given domain of interest? Should it be done through incorporating prior knowledge following the Gestalt principles (such as symmetry, proximity, and continuity)? Or should it be done through learning as infants seem to do [**FisAsl2002??**]? <ref:??>
- Lastly, the models described here focus mainly on the representation of shape, and while each is validated using a recognition of classification mechanism, that important stage of processing remains to be more carefully examined, especially in a probabilistic context. With respect to recognizing 3D surfaces embedded in images, a natural extension would be to explore Markov Random Fields or Deep Learning as computational frameworks for recognition.

In answering these questions, the main inspiration, as was true with Blum's work, will remain the same: the belief that by understanding human visual processing better we may develop better computer vision methods.

Acknowledgements This research was funded by the Natural Sciences and Engineering Research Council of Canada and Canada Research Chairs Program.

References

1. Attneave F (1954) Some informational aspects of visual perception. Psychol Rev 61:183–193
2. Bishop P, Kato H, Orban G (1980) Direction selective cells in complex family in cat striate cortex. J Neurophysiol 43:1266–1283
<unc>
3. Blum H (1962) An associative machine for dealing with the visual field and some of its biological implications. Air Force Cambridge Research Labs, L G Hanscom Field, Mass, Feb 1962
4. Blum H (1967) A transformation for extracting descriptors of shape. In: Wathen-Dunn W (ed) Models for the perception of speech and visual forms. MIT Press, Cambridge, pp 362–380
5. Booth M, Rolls E (1998) View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. Cereb Cortex 8(6):510–523
6. Boynton G, Hegde J (2004) Visual cortex: the continuing puzzle of area v2. Curr Biol 14(13):R523–R524
7. Brincat S, Connor C (2004) Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nat Neurosci 7(8):880–886
8. Cant JS, Goodale MA (2011) Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. J Neurosci 31(22):8248–8258

9. Clements DH, Sarama J (2000) What do Children Know about Shapes? In: Teaching children mathematics. April 2000, The National Council of Teachers of Mathematics, Inc, pp 482–488
10. Corbetta M, Miezin F, Dobmeyer S, Shulman GL, Petersen SE (1991) Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. J Neurosci 11(9):2393–2402
11. Dobbins A (1992) Difference models of visual cortical Neurons. Doctoral dissertation, Department of Electrical Engineering, McGill University
12. Dudek G, Tsotsos JK (1997) Shape representation and recognition from multiscale curvature. Comput Vis Image Underst 68(2):170–189
13. Dudek G, Tsotsos JK (1991) Shape representation and recognition from curvature. In: Proc computer vision and pattern recognition, pp 35–41
14. Dudek G, Tsotsos JK (1990) Recognizing planar curves using curvature-tuned smoothing. In: Proceedings 10th international conference on pattern recognition, vol 1, pp 130–135
15. Fergus R, Perona P, Zisserman A (2003) Object class recognition by unsupervised scale-invariant learning. In: CVPR, vol 2, p 264
16. Fukushima K (1980) Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36(4):193–202
17. Gershfokk-Stowe L, Smith LB (2004) Shape and the first hundred nouns. Child Dev 75(4):1098–1114
18. Han X, Chen Y, Ruan X (2010) Image recognition by learned linear subspace of combined bag-of-features and low-level features. In: ICIP
19. Hawken M, Parker A (1987) Spatial properties of neurons in the monkey striate cortex. Proc R Soc Lond B, Biol Sci 231:251–288
20. Hopf J-M, Boehler CN, Schoenfeld MA, Heinze H-J, Tsotsos JK (2010) The spatial profile of the focus of attention in visual search: insights from MEG recordings. Vis Res 50(14):1312–1320
21. Hubel D, Wiesel T (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106–154
22. Hubel D, Wiesel T (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195(1):215–243
23. Hummel JE, Stankiewicz BJ (1998) Two roles for attention in shape perception: a structural description model of visual scrutiny. Vis Cogn 5:49–79
24. Ito M, Komatsu H (2004) Representation of angles embedded within contour stimuli in area v2 of macaque monkeys. J Neurosci 24(13):3313–3324
25. James TW, Stevenson RA, Kim S, VanDerKlok RM, James KH (2011) Shape from sound: evidence for a shape operator in the lateral occipital cortex. Neuropsychologia 49:1807–1815
26. Jones SS, Smith LB (1993) The place of perception in children's concepts. Cogn Dev 8:113–139
27. Kato H, Bishop P, Orban G (1978) Hypercomplex and simple/complex cells classifications in cat striate cortex. J Neurophysiol 41:1071–1095
28. Koenderink JJ, van Doorn AJ (1980) Photomettric invariants related to solid shape. Opt Acta 27(7):981–996
29. Kruijne W, Tsotsos JK (2011) Visuo-cognitive routines: reinterpreting the theory of visual routines as a framework for visual cognition. Technical Report CSE-2011-05, Dept of Computer Science & Engineering, York University
30. Leung B (2004) Component-based car detection in street scene IMages. PhD thesis, Massachusetts Institute of Technology, Dept of Electrical Engineering and Computer Science
31. Marcelja S (1980) Mathematical description of the responses of simple cortical cells. J Opt Soc Am 70(11):1297–1300
32. Merigan W, Pham H (1998) 4 lesions in macaques affect both single and multiple-viewpoint shape discriminations. Vis Neurosci 15:359–367
33. Orban G, Kato H, Bishop P (1979) Dimensions and properties of end-zone inhibitory areas of hypercomplex cells in cat striate cortex. J Neurophysiol 42:833–849

34. Orban G, Kato H, Bishop P (1979) End-zone region in receptive fields of hypercomplex and other striate neurons in the cat. J Neurophysiol 42:818–832
35. Pasupathy A, Connor C (1999) Responses to contour features in macaque area V4. J Neurophysiol 82(5):2490–2502
36. Pasupathy A, Connor C (2001) Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophysiol 86(5):2505–2519
37. Pasupathy A, Connor C (2002) Population coding of shape in area V4. Nat Neurosci 5(12):1332–1338
38. Rektorys K (1980) Variational methods in mathematics, science and engineering. Reidel, Dordrecht
39. Riesenhuber M, Poggio T (1999) Hierarchical models of object recognition in cortex. Nat Neurosci 2(11):1019–1025
40. Rodríguez-Sánchez AJ, Simine E, Tsotsos JK (2007) Attention and visual search. Int J Neural Syst 17(4):275–288
41. Rodríguez-Sánchez A (2010) Intermediate visual representations for attentive recognition systems. PhD, York University
42. Rodriguez-Sanchez A, Tsotsos JK (2011) The importance of intermediate representations for the modeling of 2D shape detection: endstopping and curvature tuned computations. In: Proc IEEE computer vision and pattern recognition, Colorado Springs, CO
43. Rodríguez-Sánchez A, Tsotsos J (2012) The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. PLoS ONE 7(8):1–13
44. Samuelson LK, Smith LB (2005) They call it like they see it: spontaneous naming and attention to shape. Dev Sci 8(2):182–198
45. Sereno AB, Amador SC (2006) Attention and memory-related responses of neurons in the lateral intraparietal area during spatial and shape-delayed match-to-sample tasks. J Neurophysiol 95:1078–1098
46. Serre T, Wolf L, Bileschi S, Riesenhuber M (2007) Robust object recognition with cortex-like mechanisms. IEEE Trans Pattern Anal Mach Intell 29(3):411–426
47. Serre T, Wolf L, Poggio T (2005) Object recognition with features inspired by visual cortex. In: IEEE conference on computer vision and pattern recognition
48. Siddiqi K, Shokoufandeh A, Dickinson SJ, Zucker SW (1999) Shock graphs and shape matching. Int J Comput Vis 35(1):13–32
49. Sigurdardottir HM, Michalak SM, Sheinberg DL (2012) Shape beyond recognition: how object form biases spatial attention and motion perception. J Vis 12(9):665
50. Smith LB, Jones SS, Landau B, Gershkoff-Stowe L, Samuelson L (2002) Object name learning provides on-the-job training for attention. Psychol Sci 13(1):13–19
51. Spelke E (2000) Principles of object perception. Cogn Sci 14:29–56
52. Tanaka K (1996) Representation of visual features of objects in the inferotemporal cortex. Neural Netw 9(8):1459–1475
53. Terzopoulos D (1986) Regularization of inverse visual problems involving discontinuities. IEEE Trans Pattern Anal Mach Intell 8(4):413–424
54. Todd JT (2004) The visual perception of 3D shape. Trends Cogn Sci 8(3):115–121
55. Tsotsos JK (2011) A computational perspective on visual attention. MIT Press, Cambridge
56. Ullman S (1984) Visual routines. Cognition 18(1–3):97–159
57. von der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. Science 224(4654):1260–1262