Attending to Motion: Localizing and Classifying Motion Patterns in Image Sequences

John K. Tsotsos[1], Marc Pomplun[2], Yueju Liu[1], Julio C. Martinez-Trujillo[1], and Evgueni Simine[1]

[1]Centre for Vision Research, York University, Toronto, Canada M3J 1P3
[2]Department of Computer Science, University of Massachusetts at Boston, Boston, MA 02125, USA

Abstract. The Selective Tuning Model is a proposal for modelling visual attention in primates and humans. Although supported by significant biological evidence, it is not without its weaknesses. The main one addressed by this paper is that the levels of representation on which it was previously demonstrated (spatial Gaussian pyramids) were not biologically plausible. The motion domain was chosen because enough is known about motion processing to enable a reasonable attempt at defining the feedforward pyramid. The effort is unique because it seems that no past model presents a motion hierarchy plus attention to motion. We propose a neurally-inspired model of the primate visual motion system attempting to explain how a hierarchical feedforward network consisting of layers representing cortical areas V1, MT, MST, and 7a detects and classifies different kinds of motion patterns. The STM model is then integrated into this hierarchy demonstrating that successfully attending to motion patterns, results in localization and labelling of those patterns.

1 Introduction

Attentive processing is a largely unexplored dimension in the computational motion field. No matter how sophisticated the methods become for extracting motion information from image sequences, it will not be possible to achieve the goal of human-like performance without integrating the optimization of processing that attention provides. Virtually all past surveys of computational models of motion processing completely ignore attention. However, the concept has crept into work over the years in a variety of ways.

One can survey the current computer vision literature and realize that attentive processing is not much of a concern. Many recent reviews of various aspects of motion understanding have not made any mention of attentive processing of any kind [1, 2, 3, 4, 5]. The review by Aggarwal and Cai [6] includes one example of work that uses motion cues to segment an object and to affix an attentional window on to it. This is a data-directed attentional tool. Gavrila's review [7] includes one example of where vision can provide an attentional cue for speech localization. Most

of these cited papers make the claim that little or no work had been done on the topic of high level motion understanding previously (see [8] for a review that refutes this).

Many authors do not consider attention simply because of assumptions that eliminate the issue. An example of the kinds of assumptions that are typically made even in the best work follows [9]. The input to this system must satisfy the following: a) all frames in a given movie must contain the same number of figures; b) the figures in each frame must be placed in a one-to-one correspondence with figures in adjacent frames; and, c) the system must be given this correspondence as input. Others, such as in [10], assume that their algorithm starts off by being given the region of interest that corresponds to each object that may be moving. The processing that ensues is perhaps the best of its kind currently, but the algorithm critically depends on reasonable regions of interest and is not designed to find that region of interest either independently or concurrently as it processes the events in the scene. In a third example the values for the sensors are manually extracted by watching a video of the action and further, even determine the interval where every action and sub-action occurs [11]. The problem is not that any one effort makes these assumptions; the problem lies in the fact that it is now almost universal to assume the unreasonable. We are not trying to be critical of these authors; rather, the correct conclusion to draw from these comments is that we suggest a more balanced approach to the problem across the discipline, where at least some researchers study the attentive issues involved in a more general solution.

Attentive components have been included in systems not only through assumptions. At least three tools have appeared: the detection of salient tracking points/structures; search region predictions; and, Kalman filters and their extensions. Many examples have appeared [12, 13, 14, 15]. All are clearly strategies that help reduce search however, the overall result is an ad hoc collection of domain-specific methods.

A similar survey of computational neuroscience literature reveals many interesting motion models and better interest in motion attention. More discussion on these efforts appears later.

2. The Selective Tuning Model

Complexity analysis leads to the conclusion that attention must tune the visual processing architecture to permit task-directed processing [16]. In its original definition, [16], the Selective Tuning Model (STM), selection takes two forms: *spatial* selection is realized by inhibiting task-irrelevant locations in the neural network, and *feature* selection is realized by inhibiting the neurons that represent task-irrelevant features. When task constraints are available they are used to set priorities for selection; if not available, then there are default priorities (such as 'strongest response'). The two cornerstones of spatial and feature selection have since been experimentally supported [17, 18]. Only a brief summary is presented here since the model is detailed elsewhere [19].

The spatial role of attention in the image domain is to localize a subset of the input image and its path through the processing hierarchy such as to minimize any interfering or corrupting signals. The visual processing architecture is a pyramidal

network composed of units receiving both feed-forward and feedback connections. When a stimulus is first applied to the input layer of the pyramid, it activates in a feed-forward manner all of the units within the pyramid to which it is connected. The result is the activation of an inverted sub-pyramid of units and we assume that the degree of unit activation reflects the goodness-of-match between the unit and the stimulus it represents.

Attentional selection relies on a hierarchy of winner-take-all (WTA) processes. WTA is a parallel algorithm for finding the maximum value in a set of variables, which was first proposed in this context by Koch and Ullman [20]. WTA can be steered to favor particular stimulus locations or features but in the absence of such guidance it operates independently. The processing of a visual input involves three main stages. During the first stage, a stimulus is applied to the input layer and activity propagates along feed-forward connections towards the output layer. The response of each unit depends on its particular selectivities, and perhaps also on a top-down bias for task-relevant qualities. During the second stage, a hierarchy of WTA processes is applied in a top-down, coarse-to-fine manner. The first WTA process operates in the top layer and covers the entire visual field at the top layer: it computes the unit or groups of contiguous units with the largest response in the output layer, that is, the *global winner*. In turn, the global winner activates a WTA amongst its input units in the layer immediately below. This localizes the largest response within the receptive field of the global winner. All of the connections of the visual pyramid that do not contribute to the winner are pruned (i.e., attenuated). This strategy of finding the winner within each receptive fields and then pruning away irrelevant connections, is applied recursively through the pyramid, layer by layer. Thus, the global winner in the output layer is eventually traced back to its perceptual origin in the input layer. The connections that remain (i.e., are not pruned) may be considered the *pass zone* of the attentional beam, while the pruned connections an *inhibitory zone* around that beam. A final feedforward pass then allows the selected stimulus to be processed by the network without signal interference from surrounding stimuli. This constitutes a single attentive processing cycle.

The processing exhibits serial search for displays with multiple objects using a simple inhibition of return mechanism, that is, the pass zone pathways are inhibited for one processing cycle so that in the next feedforward pass the second strongest responses form the global winner and the WTA hierarchy focuses in on the second strongest item in the display. The processing operates continuously in this manner

The selective tuning model was developed with the dual goals of computational utility and biological predictive power. The predictions (appearing mostly in [16, 19]) and supporting evidence are briefly described.

• An early prediction was that attention is necessary at any level of processing where a many-to-one mapping between neural processes is found. Further, attention occurs in all the areas in a coordinated manner. The prediction was made at a time when good evidence for attentional modulation was known for area V4 only [21]. Since then, attentional modulation has been found in many other areas both earlier and later in the visual processing stream, and that it occurs in these areas simultaneously [22]. Vanduffel et al. [23] have shown that attentional modulation appears as early as the LGN. The prediction that attention modulates all cortical and even subcortical levels of processing has been borne out by recent work from several groups [23, 24, 25].

• The notions of competition between stimuli and of attentional modulation of this competition were also early components of the model and these too have gained substantial support over the years [17, 22, 27].

• The model predicts an inhibitory surround that impairs perception around the focus of attention a prediction that seems to be gaining support, both psychophysically and neurophysiologically [23, 26, 28, 29, 30, 31].

• A final prediction is that latency of attentional modulations *decreases* from lower to higher visual areas. Although controversial, it seems that attentional effects do not appear until 150 ms after the onset of a stimulus in IT cortex [32] while in V1 they appear after 230 ms [33].

Additional predictions of the selective tuning model concern the form of spatial and temporal modulations of visual cortical responses around the focus of attention, and the existence of a WTA circuit connecting cortical columns of similar selectivity. The selective tuning model offers a principled solution to the fundamental problems of visual complexity, a detailed perceptual account of both the guidance and the consequences of visual attention, and a neurally plausible implementation as an integral part of the visual cortical hierarchy. Thus, the model "works" at three distinct levels - computational, perceptual, and neural - and offers a more concrete account, and far more specific predictions, than previous models limited to one of these levels.

Previous demonstrations of the Selective Tuning Model were not without their weaknesses. The main one addressed by this paper is that the levels of representation shown in [19] were not biologically plausible. Here, the motion domain is chosen in order to demonstrate that STM can indeed operate as desired with realistic representations because enough is known about motion processing to enable a reasonable attempt at defining the feedforward pyramid. In addition, the effort is unique because it seems that no past model presented a motion hierarchy plus attention to motion [34, 35, 36, 37, 38, 39, 40, 41, 42]. The remainder of this paper will focus on this issue.

3 The Feedforward Motion Pyramid

We propose a neurally-inspired model of the primate motion processing hierarchy. The model aims to explain how a hierarchical feed-forward network consisting of neurons in the cortical areas V1, MT, MST, and 7a of primates detects and classifies different kinds of motion patterns. At best, the motion model is a first-order one with much elaboration left for future work. Indeed, some of the previous motion models offer better sophistication at one or another level of processing; however, none cover all these levels and incorporate selective attentional processes. The primary goal is to demonstrate that the STM functions not only as previously demonstrated on Gaussian pyramids but also on a more biologically realistic representation.

Cells in *striate area V1* are selective for a particular local speed and direction of motion in at least three main speed ranges [43]. In the model, V1 neurons estimate local speed and direction in five-frame, 256x256 pixel image sequences using spatiotemporal filters (e.g., [44])[1]. Their direction selectivity is restricted to 12

[1] The choices of parameters for sizes of representations, filters, etc. are mostly for convenience and variations in them has no effect on overall results intended by this demonstration.

distinct, Gaussian-shaped tuning curves. Each tuning curve has a standard deviation of 30º and represents the selectivity for one of 12 different directions spaced 30º apart (0º, 30º, …, 330º). V1 is represented by a 60x60 array of hypercolumns. The receptive fields (RFs) of V1 neurons are circular and homogeneously distributed across the visual field, with RFs of neighboring hypercolumns overlapping by 20%.

In *area MT* a high proportion of cells are tuned for a particular local speed and direction of movement, similar to direction and speed selective cells in V1 [45, 46]. A proportion of MT neurons are also selective for a particular angle between movement direction and spatial speed gradient [47]. Both types of neurons are represented in the MT layer of the model, which is a 30x30 array of hypercolumns. Each MT cell receives input from a 4x4 field of V1 neurons with the same direction and speed selectivity.

Neurons in *area MST* are tuned to complex motion patterns: expand or approach, contract or recede, rotation, with RFs covering most of the visual field [48, 49]. Two types of neurons are modeled: one type selective for translation (as in V1) and another type selective for spiral motion (clockwise and counterclockwise rotation, expansion, contraction and combinations). MST is simulated as a 5x5 array of hypercolumns. Each MST cell receives input from a large group (covering 60% of the visual field) of MT neurons that respond to a particular motion/gradient angle. Any coherent motion/gradient angle indicates a particular type of spiral motion.

Finally, *area 7a* seems to involve at least four different types of computations [50]. Here, neurons are selective for translation and spiral motion as in MST, but they have even larger RFs. They are also selective for rotation (regardless of direction) and radial motion (regardless of direction). In the simulation, area 7a is represented by a 4x4 array of hypercolumns. Each 7a cell receives input from a 4x4 field of MST neurons that have the relevant tuning. Rotation cells and radial motion cells only receive input from MST neurons that respond to spiral motion involving any rotation or any radial motion, respectively.

Fig. 1 shows the resulting set of neural selectivities that comprise the entire pyramidal hierarchy covering visual areas V1, MT, MST and 7a. It bears repeating that this should only be considered a first order model.

Fig. 2 shows the activation of neurons in the model as induced by a sample stimulus. Note that in the actual visualization different colors indicate the response to particular angles between motion and speed gradient in MT gradient neurons. In the present example, the gray levels indicate that the neurons selective for a 90º angle gave by far the strongest responses. A consistent 90º angle across all directions of motion signifies a pattern of clockwise rotation. Correspondingly, the maximum activation of the spiral neurons in areas MST and 7a corresponds to the clockwise rotation pattern (90º angle). Finally, area 7a also shows a substantial response to rotation in the medium-speed range, while there is no visible activation that would indicate radial motion.

Figures 3, 4, 5 and 6 provide additional detail required for explanation of Figures 1 and 2.

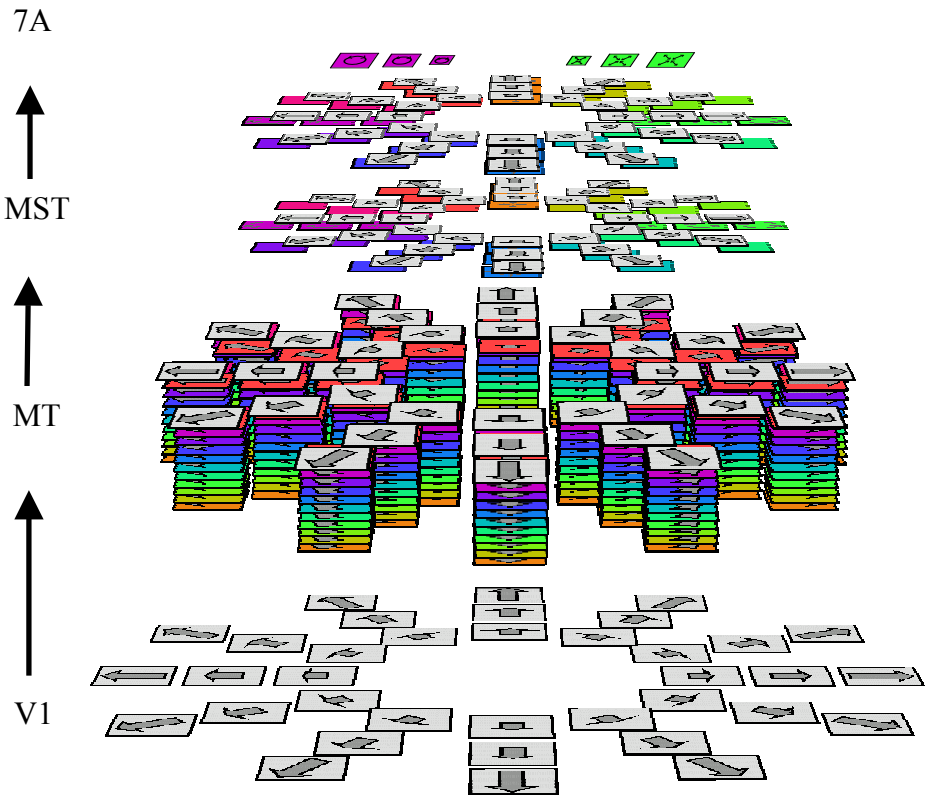

Fig. 1. The overall set of representations for the different types of neurons in areas V1, MT, MST, and 7a. Each rectangle represents a single type of selectivity applied over the full image at that level of the pyramid. Large grey arrows represent selectivity for direction. Coloured rectangels represent particular angles between motion and speed gradient. The three rectangles at each direction represent the three speed selectivity ranges in the model. In this way, each single 'sheet' may be considered an expanded view of the 'hypercolumns' in a visual area. In area V1, for example, direction and speed selectivities are represented by the single sheet of rectangles in the figure. In area MT, there are 13 sheets, the top one representing direction and speed selectivity while the remaining 12 represent the 12 directions of speed gradient for each combination of speed and direction ranges (Fig. 4 provides additional explanation of the speed gradient coding). MST units respond to patterns of motion – contract, recede, and rotate. This figure emphasizes the scale of the search problem faced by the visual system: to determine which responses within each of these representations belong to the same event.

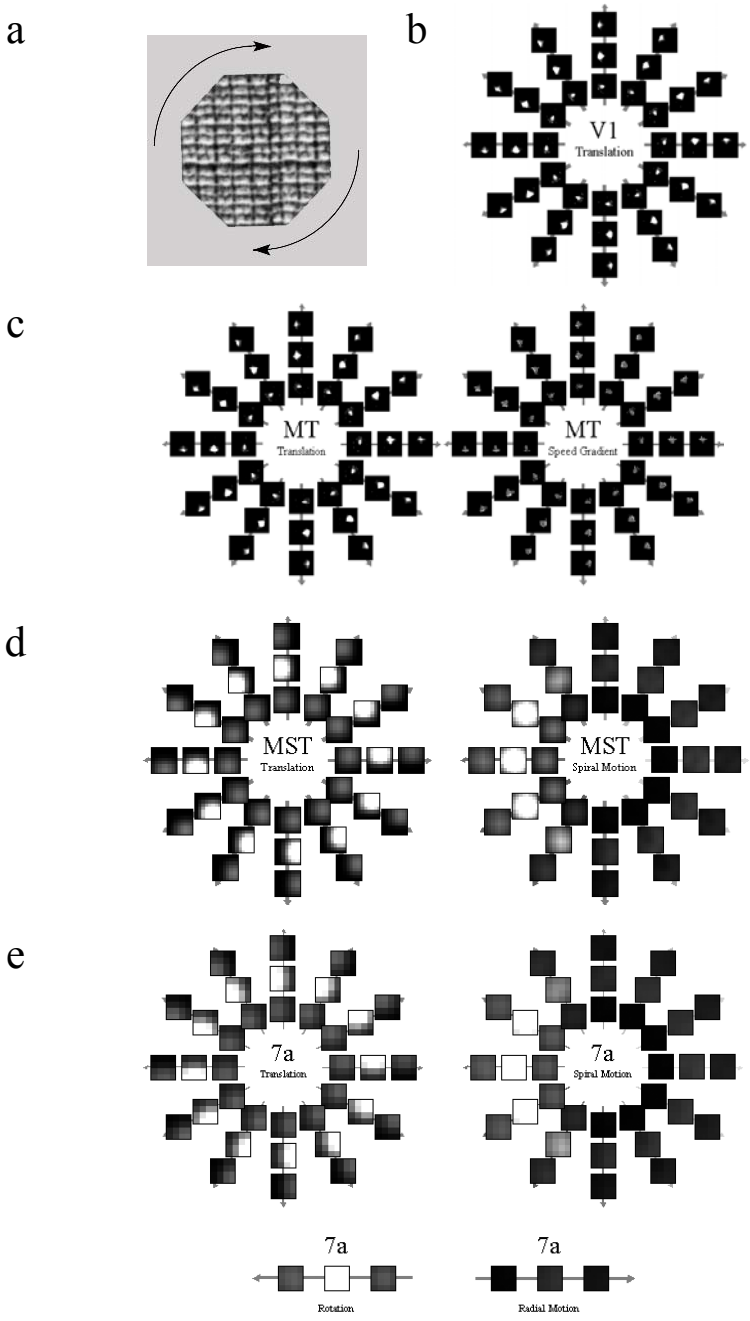

Fig. 2. The model's response to a clockwise rotating stimulus (a). Brightness indicates activation in areas V1, MT, MST, and 7a (b to e). Each of the figures represents the output of one representational sheet as depicted in Fig. 1. As is clear, even with a single object undergoing a single, simple motion, a large number of neurons respond.

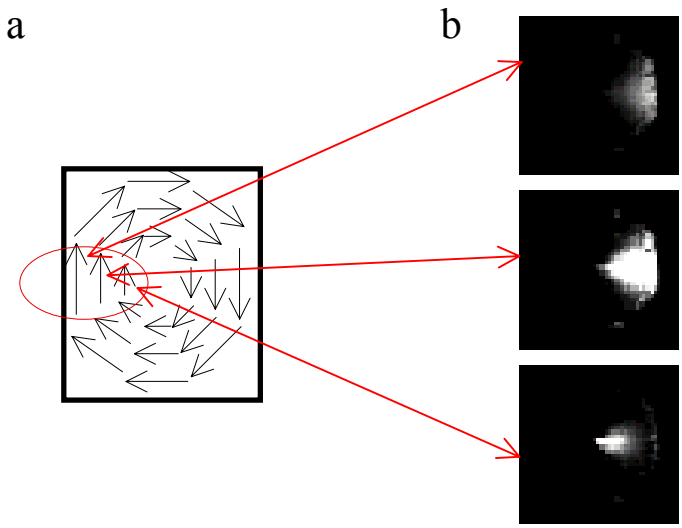

Fig. 3. Detail from area V1 in Fig. 2. (a) A depiiction of the optic flow vectors resulting from the rotating motion. (b) The three speed selectivities for 'upwards' direction selectivity, the top being fast, the middle medium and the bottom low speed. The brightness shows responses across the sub-image due to the motion.

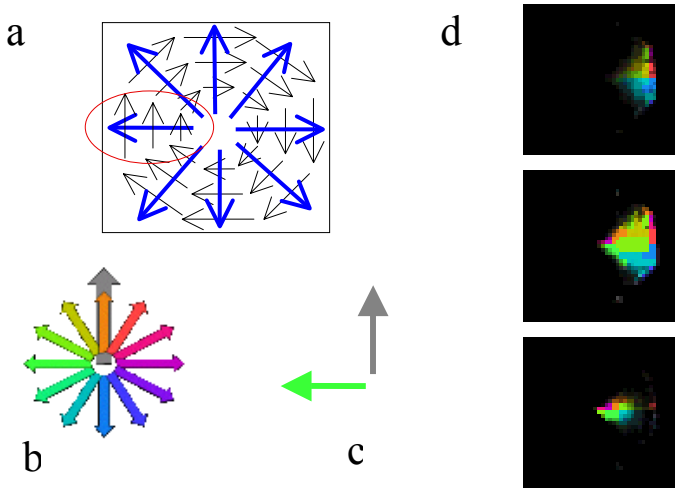

Fig. 4. Detail from area MT in Fig. 2. (a) The direction of the speed gradient for the rotating optic flow is shown with blue arrows. The red oval shows the only portion of the stimulus that activates the vertical motion selectivite neurons shown in (d), similarly to Fig. 3. (b) The colour coding used for the different directions of speed gradient relative to the direction of motion given by the gray arrow. (c) The particular 'ideal' speed gradient/direction tuning for the stimuli within the red oval. (d) Responses of the MT neurons with the tuning in (c) for three different speeds, the top being fast. The stimulus is the one shown in Fig. 2; responses are not perfectly clean (i.e., all light green in colour) due to the noise inherent in the processing stages.

Fig. 5. Detail from Fig. 2 for the neurons representing motion patterns in area MST. As is clear, the 'brightest' (strongest) responses occur in the representation of medium speed, clockwise rotation. There are many other responses some rather strong, through the sheet. It is the task of attentional selection to determine which responses are the correct ones to focus on in order to optimally localize the stimulus.

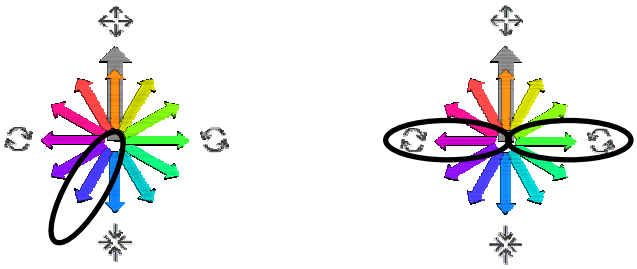

Fig. 6. Two examples of speed gradient coding. (a) If the stimulus object is both rotating clockwise and receding, the responses in area MT are coded blue. (b) If there are two objects in the image one rotating clockwise and the other counterclockwise, the responses in area MT will be coded light purple for the spatial extent of the former and light green for the spatial extent of the latter. Neurons in area MST spatially group common MT responses. The attention system then segments one from the other based on strength of response and motion type.

4 Using STM to Attend to and Localize Motion Patterns

Most of the computational models of primate motion perception that have been proposed concentrate on feedforward, classical types of processing and do not address

attentional issues. However, there is strong evidence that the responses of neurons in areas MT and MST are modulated by attention [51]. As a result of the model's feedforward computations, the neural responses in the high- level areas (MST and 7a) roughly indicate the kind of motion patterns presented as an input but do not localize the spatial position of the patterns. The STM model was then applied to this feedforward pyramid, adding in the required feedback connections, hierarchical WTA processes, and gating networks as originally defined in [16, 19]. The result is that the model attends to object motion, whether it exhibits a single or concurrent motion, and serially focuses on each motion in the sequence in order of response strength.

The integration of the STM into this feedforward network requires one additional component not previously described. A *motion activity map* with the same size as a 7a layer is constructed after the feedforward processing. The value of a node in the activity map is a weighted sum of the activations of all 7a neurons at this position and it reflects the overall activation across all motion patterns. A location-based weighted sum is required in order to correctly detect single objects exhibiting simultaneous multiple motion types. This is not the same as the saliency map of [20] since it is not based on point locations and does not solely determine the attended region. Second, the hierarchical described earlier finds the globally most active region. Then for this region, two separate WTAs compete among all the translational motion patterns and spiral motion patterns respectively and thus result in a winning region is each representation. The remainder of processing proceeds as described in Section 2.0 for each of the winning patterns. Although not described here, the model also includes processes for tracking translating objects and for detecting onset and offset events (start and stop). Figures 7 and 8 present a 3D visualization of the model receiving an image sequence that contains an approaching object and a counterclockwise rotating object.

Fig. 7. The first image of the sequence used as demonstration in the next figure. The checkerboard is rotating while the box (in one of the authors' hands) is approaching the camera.

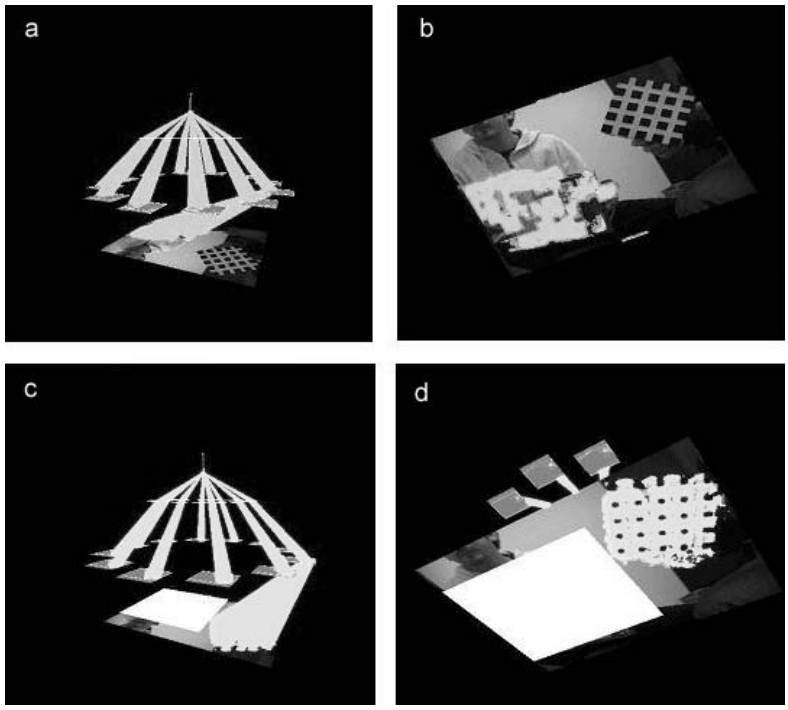

Fig. 8. Visualization of the attentional mechanism applied to an image sequence showing an approaching object and a counterclockwise rotating object at the same time. First, the model detects the approaching motion and attends to it (a); the localization of the approaching object can be seen most clearly from below the motion hierarchy (bright area in panel b). Then, the pass zone associated with it is inhibited, and the model attends to the rotating motion (c and d).

5. Discussion

Due to the incorporation of functionally diverse neurons in the motion hierarchy, the output of the present model encompasses a wide variety of selectivities at different resolutions. This enables the computer simulation of the model to detect and classify various motion patterns in artificial and natural image sequences showing one or more moving objects as well as single objects undergoing complex, multiple motions. Most other models of biological motion perception focus on a single cortical area. For instance, the models by Simoncelli and Heeger [34] and Beardsley and Vaina [35] are biologically relevant approaches that explain some specific functionality of MT and MST neurons, respectively, but do not include the embedding hierarchy in the motion pathway. On the other hand, there are hierarchical models for the detection of motion (e.g., [36, 37]), but unlike the present model they do not provide a biologically plausible version of the motion processing hierarchy.

Another strength of our model is its mechanism of visual attention. To our knowledge, there are only 2 other motion models employing attention for motion. The earlier one is due to Nowlan and Sejnowski [38]. There, processing that is much in the

same spirit as ours but very different in form takes place. They compute motion energy with the goal of modelling MT neurons. This energy is part of a hierarchy of processes that include softmax for local velocity selection. They suggest that the selection permits processing to be focussed on the most reliable estimates of velocity. There is no top-down component nor full processing hierarchy. The relationship to attentional modulation that has been described after their model was presented of course is not developed; it does not appear to be within the scope of their model. The second one is from Grossberg, Mingolla, and Viswanathan [39], which is a motion integration and segmentation model for motion capture. Called the Formotion BCS model, their goal is to integrate motion information across the image and segment motion cues into a unified global percept. They employ models of translational processing in areas V1, V2, MT and MST and do not consider motion patterns. Competition determines local winners among neural responses and the MST cells encoding the winning direction have an excitatory influence on MT cells tuned to the same direction. A variety of motion illusions are illustrated but no real image sequences are attempted. Neither model has the breadth of processing in the motion domain or in attentional selection as the current work.

Of course, this is only the beginning and we actively pursuing several avenues of further work. The tuning characteristics of each of the neurons only coarsely model current knowledge of primate vision. The model includes no cooperative nor competitive processing among units within a layer. Experimental work examining the relationship of this particular structure to human vision is also on-going

Acknowledgements

We thank Albert Rothenstein for providing valuable comments on drafts of this paper. The work is supported by grants to JKT from the Natural Sciences and Engineering Research Council of Canada and the Institute for Robotics and Intelligent Systems, one of the Government of Canada Networks of Centres of Excellence.

References

1. Aggarwal, J.K., Cai, Q., Liao, W., Sabata, B. (1998). Nonrigid motion analysis: Articulated and elastic motion, *Computer Vision and Image Understanding 70(2)*, p142–156.
2. Shah, M., Jain, R. (1997). Visual recognition of activities, gestures, facial expressions and speech: an introduction and a perspective, in **Motion-Based Recognition**, ed. by M. Shah and R. Jain, Kluwer Academic Publishers.
3. Cedras, C., Shah, M. (1994). A survey of motion analysis from moving light displays, IEEE CVPR-94, Seattle, Washington, p214-221.
4. Cedras, C. , Shah, M. (1995). Motion-based recognition: A survey, *Image and Vision Computing, 13(2)*, p129-155.
5. Hildreth, E. Royden, C. (1995). Motion Perception, **in The Handbook of Brain Theory and Neural Networks**, ed. by M. Arbib, MIT Press, p585 – 588.
6. Aggarwal , J.K., Cai, Q. (1999). Human motion analysis: A Review, *Computer Vision and Image Understanding 73(3)*, p428–440.

7. Gavrila, D.M. (1999). The visual analysis of human movement: A Survey, *Computer Vision and Image Understanding 73(1*), p82–98.
8. Tsotsos, J.K., (2001). Motion Understanding:Task-Directed Attention and Representations that link Perception with Action, Int. J. of Computer Vision 45:3, 265-280.
9. Siskind, J. M. (1995). Grounding Language in Perception. *Artificial Intelligence Review 8*, p371 - 391.
10. Mann, R., Jepson, A., Siskind, J. (1997). The computational perception of scene dynamics, *Computer Vision and Image Understanding, 65(2)*, p113 - 128.
11. Pinhanez, C., Bobick, A. (1997). Human action detection using PNF propagation of temporal constraints, MIT Media Lab TR 423, April.
12. Tsotsos, J.K. (1980). A framework for visual motion understanding, Ph.D. Thesis, Dept. of Computer Science, University of Toronto, May.
13. Dickmanns, E.D., Wünsche, H.J. (1999). Dynamic vision for perception and control of motion, **Handbook of Computer Vision and Applications Vol. 2**, ed by B. Jahne, H. Haubeccker, P. Geibler, Academic Press.
14. Dreschler, L., Nagel, H.H. (1982). On the selection of critical points and local curvature extrema of region boundaries for interframe matching, Proc. Int. Conf. Pattern Recognition, Munich, p542-544.
15. Wachter, S., Nagel, H.H. (1999). Tracking persons in monocular image sequences, *Computer Vision and Image Understanding 74(3)*, p174–192.
16. Tsotsos, J.K. (1990). Analyzing vision at the complexity level, *Behavioral and Brain Sciences 13-3*, p423 - 445.
17. Desimone, R., Duncan, J., (1995). Neural Mechanisms of Selective Attention, *Annual Review of Neuroscience 18*, p193 - 222.
18. Treue, S., Martinez-Trujillo, J.C., (1999). Feature-based attention influences motion processing gain in macaque visual cortex, nayure, 399, 575 – 579.
19. Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N. & Nuflo, F. (1995). Modeling visual attention via selective tuning. Artificial Intelligence, 78, 507-545.
20. Koch, C., Ullman, S., (1985). Shifts in selective visual attention: Towards the underlying neural circuitry, *Hum. Neurobiology 4*, p219 - 227.
21. Moran, J., Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex, *Science 229,* p782 - 784.
22. Kastner, S., De Weerd, P., Desimone, R., Ungerleider, L. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI, *Science 282*, p108 - 111.
23. Vanduffel, W., Tootell, R., Orban, G. (2000). Attention-dependent suppression of metabolic activity in the early stages of the macaque visual system, *Cerebral Cortex 10,* p109-126.
24. Brefczynski J.A., DeYoe E.A. (1999). A physiological correlate of the 'spotlight' of visual attention. *Nat Neurosci*. Apr;2(4), p370-374 .
25. Gandhi S.P., Heeger D.J., Boynton G.M. (1999). Spatial attention affects brain activity in human primary visual cortex, *Proc Natl Acad Sci U S A*, Mar 16;96(6), p3314-9 .
26. Smith, A., Singh, K., Greenlee, M. (2000). Attentional suppression of activity in the human visual cortex, *NeuroReport*, Vol 11 No 2 7, p271 – 277.
27. Reynolds, J., Chelazzi, L., Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4, *The Journal of Neuroscience, 19(5)*, p1736–1753.
28. Caputo, G., Guerra, S. (1998). Attentional selection by distractor suppression, *Vision Research 38(5)*, p669 - 689.
29. Bahcall, D., Kowler, E. (1999). Attentional interference at small spatial separations, *Vision Research 39(1)*, p71 - 86.
30. Tsotsos, J.K., Culhane, S., Cutzu, F. (2001). From theoretical foundations to a hierarchical circuit for selective attention, **Visual Attention and Cortical Circuits**, ed. by J. Braun, C. Koch and J. Davis, p285 – 306, MIT Press.

31. Cutzu, F., Tsotsos, J.K., The selective tuning model of visual attention: Testing the predictions arisiing from the inhibitory surround mechanism, Vision Research, (in press)
32. Chelazzi, L., Duncan, J., Miller, E., Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search, *J. Neurophysiology 80*, p2918 - 2940.
33. Roelfsema, P., Lamme, V., Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey, *Nature 395*, p376 - 380.
34. Simoncelli, E.P. & Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Research, 38 (5), 743-761.
35. Beardsley, S.A. & Vaina, L.M. (1998). Computational modeling of optic flow selectivity in MSTd neurons. Network: Computation in Neural Systems, 9, 467-493.
36. Giese, M.A. (2000). Neural field model for the recognition of biological motion. Paper presented at the Second International ICSC Symposium on Neural Computation (NC 2000), Berlin, Germany.
37. Meese, T.S. & Anderson, S.J. (2002). Spiral mechanisms are required to account for summation of complex motion components. Vision Research, 42, 1073-1080.
38. Nowlan, S.J., Sejnowski, T.J., (1995). A Selection Model for Motion Processing in Area MT of Primates, The Journal of Neuroscience 15 (2), p 1195 – 1214.
39. Grossberg, S., Mingolla, E. & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. Vision Research, 41, 2521-2553.
40. Zemel, R. S., Sejnowski, T.J., (1998). A Model for Encoding Multiple Object Motions and Self-Motion in area MST of Primate visual cortex, The Journal of Neuroscience, 18(1), 531 – 547.
41. Pack, C., Grossberg, S. Mingolla, E., (2001). A nerual model of smooth pursuit control and motion perception by cortical area MST, Journal of Cognitive Neuroscience, 13(1), 102 – 120.
42. Perrone, J.A. & Stone, L.S. (1998) Emulating the visual receptive field properties of MST neurons with a template model of heading estimation. *The Journal of Neuroscience*, 18, 5958-5975.
43. Orban, G.A., Kennedy, H. & Bullier, J. (1986). Velocity sensitivity and direction sensitivity of neurons in areas V1 and V2 of the monkey: Influence of eccentricity. Journal of Neurophysiology, 56 (2), 462-480.
44. Heeger, D.J. (1988). Optical flow using spatiotemporal filters. International Journal of Computer Vision, 1 (4), 279-302.
45. Lagae, L., Raiguel, S. & Orban, G.A. (1993). Speed and direction selectivity of Macaque middle temporal neurons. Journal of Neurophysiology, 69 (1), 19-39.
46. Felleman, D.J. & Kaas, J.H. (1984). Receptive field properties of neurons in middle temporal visual area (MT) of owl monkeys. Journal of Neurophysiology, 52, 488-513.
47. Treue, S. & Andersen, R.A. (1996). Neural responses to velocity gradients in macaque cortical area MT. Visual Neuroscience, 13, 797-804.
48. Graziano, M.S., Andersen, R.A. & Snowden, R.J. (1994). Tuning of MST neurons to spiral motions. Journal of Neuroscience, 14 (1), 54-67.
49. Duffy, C.J. & Wurtz, R.H. (1997). MST neurons respond to speed patterns in optic flow. Journal of Neuroscience, 17(8), 2839-2851.
50. Siegel, R.M. & Read, H.L. (1997). Analysis of optic flow in the monkey parietal area 7a. Cerebral Cortex, 7, 327-346
51. Treue, S. & Maunsell, J.H.R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature, 382, 539-541.