A Neurally-Inspired Model for Detecting and Localizing Simple Motion Patterns in Image Sequences

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

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

Abstract. In the present paper, we propose a neurally-inspired model of the primate motion processing hierarchy and describe its implementation as a computer simulation. The model aims to explain how a hierarchical feedforward network consisting of neurons in the cortical areas V1, MT, MST, and 7a of primates achieves the detection of different kinds of motion patterns. Moreover, the model includes a feedback gating network that implements a biologically plausible mechanism of visual attention. This mechanism is used for sequential localization and fine-grained inspection of every motion pattern detected in the visual scene.

1 The Feedforward Mechanism of Motion Detection

In the present paper, we propose a neurally-inspired model of the primate motion processing hierarchy and describe its implementation as a computer simulation. 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 achieves the detection of different kinds of motion patterns.

Cells in *striate area V1* are well known to be tuned towards a particular local speed and direction of motion in at least three main speed ranges [1]. In the model, V1 neurons estimate local speed and direction in five-frame, 256×256 pixel image sequences using spatiotemporal filters (e.g., [2]). Their direction selectivity is restricted to 12 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 60×60 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 towards a particular local speed and direction of movement, similar to direction and speed selective cells in V1 [3, 4]. A proportion of MT neurons are also selective for a particular angle between movement direction and spatial speed gradient [5]. Both types of neurons are represented in the MT layer of the model, which is a 30×30 array of hypercolumns. Each MT cell receives input from a 4×4 field of V1 neurons with the same direction and speed selectivity.

Neurons in *area MST* are tuned to complex motion patterns: expand or approach, shrink or recede, rotation, with RFs covering most of the visual field [6, 7]. 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 5×5 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 [8]. 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 4×4 array of hypercolumns. Each 7a cell receives input from a 4×4 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 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.

2 The Feedback Mechanism of Visual Attention

Most of the computational models of primate motion perception that have been proposed concentrate on bottom-up processing and do not address attentional issues. However, there is evidence that the responses of neurons in areas MT and MST can be modulated by attention (Treue & Maunsell, 1996). Moreover, we claim that attention is necessary for a precise localization of motion patterns in image sequences. 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.

In order to create a comprehensive motion model that is in agreement with biological findings and is capable of localizing motion patterns, we added a mechanism of visual attention to it. We decided to use the biologically plausible Selective Tuning approach [9], requiring the introduction of a feedback gating network to the model. Each neuron in the original motion hierarchy received an assembly of gating units that control the bottom-up information flow to that neuron.

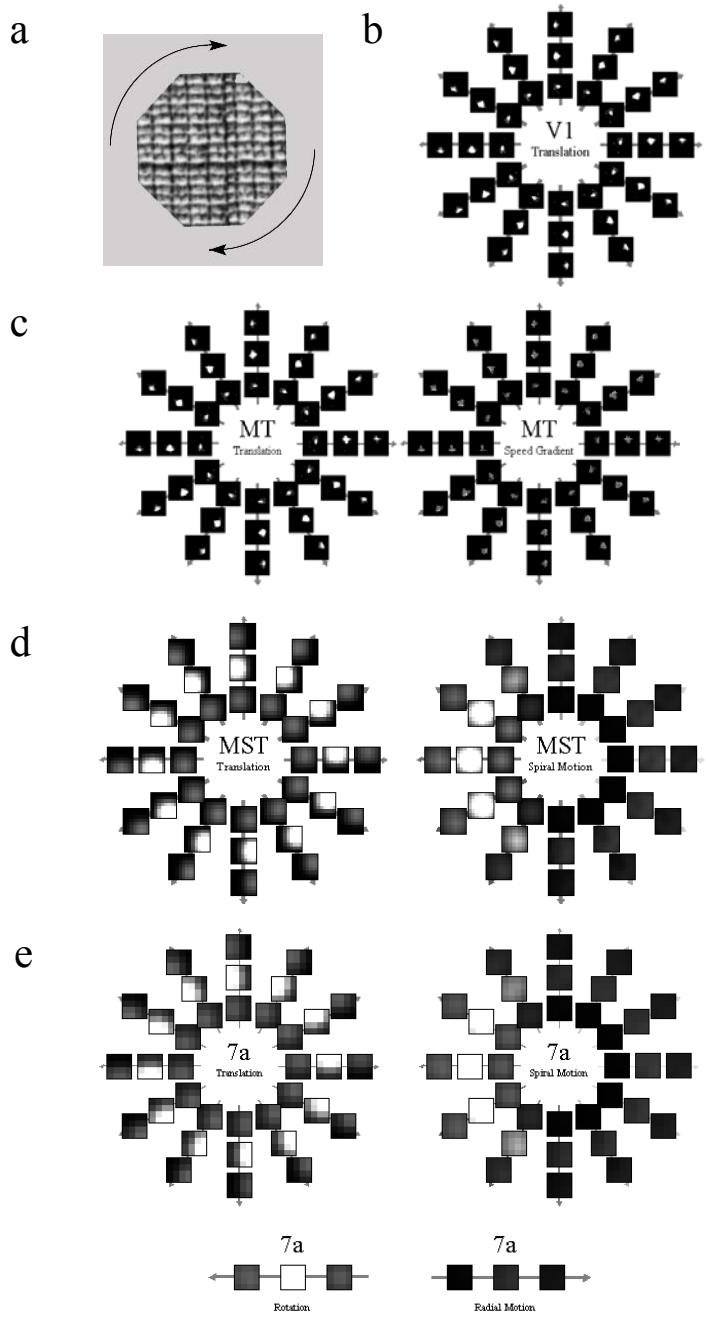

Fig. 1. The model's response to a clockwise rotating stimulus (panel a). Brightness indicates activation in areas V1, MT, MST, and 7a (panels b to e). Arrows represent selectivity for direction of motion or the angle between motion and speed gradient, and the three concentric circles stand for the three speed selectivity ranges in the model.

The attentional processing works as follows: First, a “motion activity" map with the same size as a 7a layer is constructed after the bottom-up 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. Second, a WTA (Winner-Take-All) algorithm finds the globally most active location. Then at this location, two WTAs will compete among all the translational motion patterns and spiral motion patterns respectively and thus result in two winner neurons. A WTA runs among the winners’ gating units, whose activation pattern is initially identical to the one in the winner neurons’ RFs. The resulting winners activate the connected neurons in lower layers, whereas the bottom-up information flow through the losing gating units is inhibited. This process continues until the bottom layer, and the recognized motions are localized in the input sequence. The gating network then inhibits the feed-forward processing of neighboring motion patterns so that no interfering information reaches the higher levels of the model. Loosely speaking, the model “focuses its attention“ on the winning motion pattern. Afterwards, a simple inhibition of return mechanism induces the model to switch attention to the second most active motion, and so on.

In addition, the wirings between the neurons within the same layer and the direction-selective attribute of some of the neurons enable our model to do a simplified constant motion tracking. If a neuron sensitive to motion direction a is activated at time t, then it passes its activation to neighboring neurons in the direction a at time t+1. In this way, the model focuses on the relevant area without recomputation of the whole motion hierarchy under the assumption that the motions do not change with time. In addition to tracking motion, a simple method for detecting the start and stop of motion is included. We applied a DOG operator to the area MST to detect motion changes [10]. Fig. 2 presents a 3D visualization of the model receiving an image sequence that contains an approaching object and a counterclockwise rotating object. Both motion patterns are correctly detected and localized.

3 Discussion and Conclusions

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. Most other models of biological motion perception focus on a single cortical area. For instance, the models by Simoncelli and Heeger [11] and Beardsley and Vaina [12] are biologically adequate 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., [13, 14]), but unlike the present model they do not provide a biologically plausible replica of the motion processing hierarchy in primates.

Another strength of our model is its mechanism of visual attention. To our knowledge, the only other motion model employing attention is the one by Grossberg, Mingolla, and Viswanathan [15], which is a motion integration and segmentation

model for motion capture. Their idea is that MST cells tuned to the winning direction have an excitatory influence on MT cells tuned to the same direction and nonspecifically inhibit all directionally tuned cells in MT. This kind of top-down influence from MST to MT has not been proved to exist yet. The current knowledge of effects of attention on single cell responses in area MT and MST suggests that cells in these areas have stronger responses when attention is directed into their RFs relative to when attention is directed outside the RF [16], which is compatible with our model.

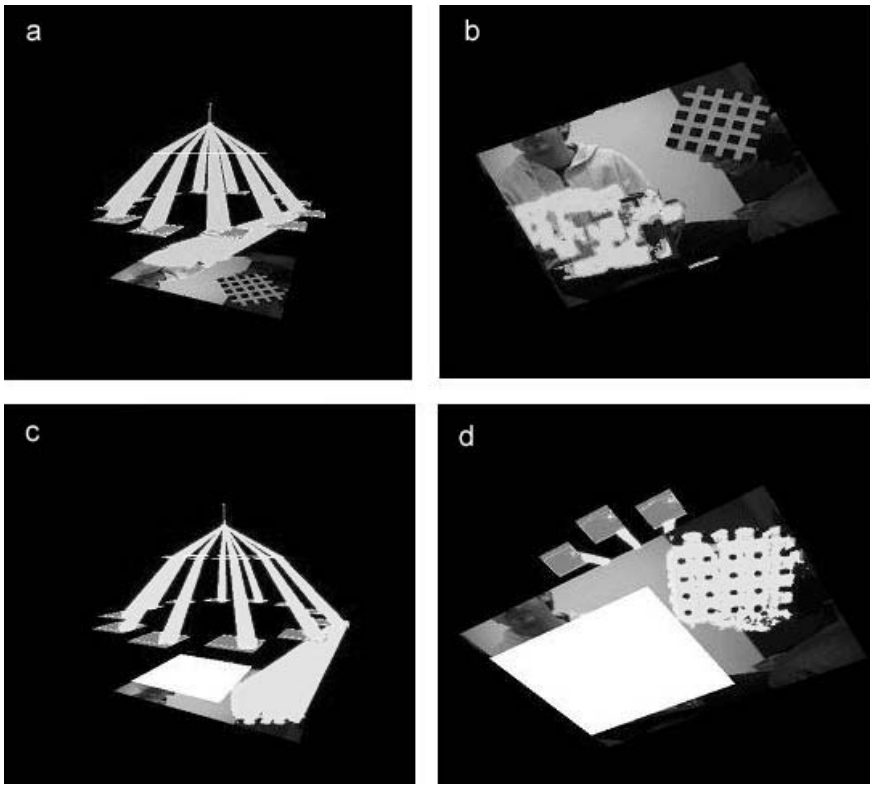

Fig. 2. Visualization of the attentional mechanism applied to an image sequence showing an approaching object and a counterclockwise rotating object at the same time. First, the model detects the approaching motion and attends to it (panel a); the localization of the approaching object can be seen most clearly from below the motion hierarchy (bright area in panel b). Then, input from the activated area is inhibited, and the model attends to the rotating motion (panels c and d).

The model has been tested on a variety of artificial and real image sequences. Simple motion patterns such as rotation, expansion, translation or combined motions with two or three patterns can be correctly recognized, localized in the image sequences and attended serially. Simple dynamic motions such as motion start, motion stop and motion pattern changes have been correctly detected as well. We

conclude that by combining four stages of motion processing with an attentional mechanism, our approach yields a biologically plausible model of visual motion processing. No current motion processing system, whether biologically inspired or not, exhibits such labeling and spatial-localization of motion patterns in image sequences.

The compatibility of our model with current neurophysiological findings and its incorporation of the diverse types of neurons found in the motion pathways provide it with predictive power for biological vision systems. Some of its predictions about activation patterns in V1, MT and MST are currently being tested in fMRI experiments on human subjects. Future work will address the perception of ego-motion, including the use of the model for controlling autonomous robots.

References

1. 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.
2. Heeger, D.J. (1988). Optical flow using spatiotemporal filters. International Journal of Computer Vision, 1 (4), 279-302.
3. Lagae, L., Raiguel, S. & Orban, G.A. (1993). Speed and direction selectivity of Macaque middle temporal neurons. Journal of Neurophysiology, 69 (1), 19-39.
4. 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.
5. Treue, S. & Andersen, R.A. (1996). Neural responses to velocity gradients in macaque cortical area MT. Visual Neuroscience, 13, 797-804.
6. Graziano, M.S., Andersen, R.A. & Snowden, R.J. (1994). Tuning of MST neurons to spiral motions. Journal of Neuroscience, 14 (1), 54-67.
7. Duffy, C.J. & Wurtz, R.H. (1997). MST neurons respond to speed patterns in optic flow. Journal of Neuroscience, 17(8), 2839-2851.
8. Siegel, R.M. & Read, H.L. (1997). Analysis of optic flow in the monkey parietal area 7a. Cerebral Cortex, 7, 327-346.
9. 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.
10. Wai, W.Y.K. (1994). A computational model for detecting image changes. Master's thesis, Department of Computer Science, University of Toronto, Ontario, Canada.
11. Simoncelli, E.P. & Heeger, D.J. (1998). A model of neuronal responses in visual area MT. Vision Research, 38 (5), 743-761.
12. Beardsley, S.A. & Vaina, L.M. (1998). Computational modeling of optic flow selectivity in MSTd neurons. Network: Computation in Neural Systems, 9, 467-493.
13. 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.
14. Meese, T.S. & Anderson, S.J. (2002). Spiral mechanisms are required to account for summation of complex motion components. Vision Research, 42, 1073-1080.
15. Grossberg, S., Mingolla, E. & Viswanathan, L. (2001). Neural dynamics of motion integration and segmentation within and across apertures. Vision Research, 41, 2521-2553.
16. Treue, S. & Maunsell, J.H.R. (1996). Attentional modulation of visual motion processing in cortical areas MT and MST. Nature, 382, 539-541.