[Submitted conference paper summary, January 1997.]

Neural model of transfer-of-binding in visual relative motion perception

Jonathan A. Marshall, Charles P. Schmitt, George J. Kalarickal, Richard K. Alley
Department of Computer Science, CB 3175, Sitterson Hall
University of North Carolina, Chapel Hill, NC 27599-3175, U.S.A.
marshall@cs.unc.edu, schmitt@cs.unc.edu, kalarick@cs.unc.edu, alley@cs.unc.edu
919-962-1887, 919-962-1881

Abstract

How can a visual system or cognitive system use the changing relationships between moving visual elements to decide which elements belong together as *groups* (or *objects*)? We have constructed a neural circuit model that selects object groupings based on global Gestalt common-fate evidence and uses information about the behavior of each group to predict the behavior of elements of the group. A simple competitive neural circuit binds elements into a representation of an object. Information about the spiking pattern of neurons allows *transfer* of the bindings of an object representation from location to location in the neural circuit as the object moves. The model exhibits characteristics of human object grouping and solves some key neural circuit design problems in visual relative motion perception.

Introduction

Human visual systems are much more sensitive to relative motion than to absolute motion. For example, the relative motion of two dots on a blank background is more easily detected than the motion of a single dot on a blank background. If both dots are also moved relative to the background, then their motion relative to each other remains more easily detected than their motion relative to the background. Each dot thus seems to provide a *reference frame* for the other's motion.

It is challenging to construct a theory of relative motion perception because such a theory must answer a difficult question: Which objects belong to which reference frames? We have designed and simulated a neural circuit model that provides an answer to this question. The model uses a simple competitive neural circuit to bind visual elements into a representation of an object. The model also uses information about the spiking pattern of neurons to *transfer* the bindings of an object representation from location to location in the neural circuit as the object moves. The model exhibits many characteristics of human perceptual relativity and solves several key neural circuit design problems: (1) the receptive fields that an object activates are selected dynamically within the context of the reference frame to which the object belongs; (2) a neural mechanism for "binding" or grouping operates to specify which parts of a scene form the reference frame for each object; (3) each object's bindings persist despite the object's motion; and (4) the groupings operate hierarchically, so that reference frames are represented relative to other reference frames.

Grouping example: Visual relative motion perception

As a conceptual example to help us design the model, we used the simple dot motion displays of Figure 1, in which correct grouping leads to improved predictions. When several dots have the same motion (Figure 1A), they are seen as constituting a group (i.e., an object, Figure 1B) *(Koffka, 1935; Ternus, 1926; Johanssen, 1950, 1985)*. When one dot in such a grouping becomes invisible (Figure 1C), it is perceived as moving *behind* an occluder *(Ramachandran, Inada, & Kiama, 1986)*. If it is represented as *continuing to exist* (Figure 1D) behind the occluder, then this representation can help the system predict correctly that it will reappear when it passes the occluder and will do so while maintaining the same relationships (e.g., distance, colinearity) to the other dots *(Marshall,* *Alley, & Hubbard, 1996).* In Figure 1E, the visible moving dots change direction while one dot is occluded. If the dots all really do belong to the same object, then the best representation is one in which the whole object changes direction (Figure 1F). What is the best prediction that the system can generate for the motion of the invisible dot? If grouping mechanisms did not exist, then the system could generate predictions based only on the motion of the dot itself, extrapolating from the dot's motion when it was visible, predicting a linear trajectory (Figure 1G). But if a grouping mechanism does exist and the dots are grouped together, then the system can predict a trajectory that follows the group motion, as indicated by the visible dots (Figure 1H). If the dots are genuinely part of the same object, then the predicted reappearance of the dot will match its actual reappearance, and this prediction will be more accurate than the predictions not based on grouping.

Figure 1: Grouping improves predictions in 3-dot motion displays. (A) Three dots with common motion are perceived as being grouped together (B) as parts of the same object. When one dot is temporarily occluded (C), it is perceived as continuing to exist (D) "behind" the occluder. How should the object's motion be represented if it changes while one dot is occluded (E)? If the three dots are indeed part of the same object (F), then a linear representation of the middle dot's motion (G) "behind" the occluder would fail. The best representation of the object's motion would be for the whole object, including the middle dot, to be represented as changing direction (H), on the basis of the two visible end dots' motion. If the middle dot has a horizontal motion relative to the group motion (J), then its reappearance from behind an occluder (I) is harder to predict. Again, a linear prediction (K) would fail. Howevever, assigning the group motion to the occluded dot would not work; instead, the group motion must be *added* to a remembered representation of the middle dot's motion, to predict the correct group-based change in the middle dot's motion (L).

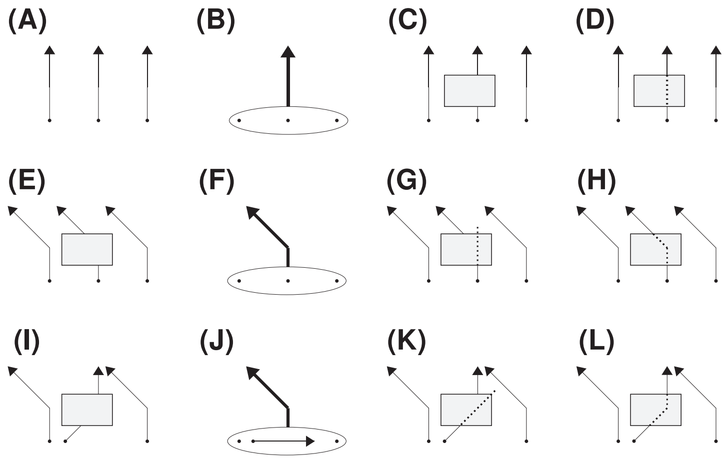


Figure 1I depicts a three-dot display in which the middle dot is seen as having a horizontal motion relative to the motion of the group of dots (Figure 1J), and in which the group motion changes direction while the middle dot is occluded. A linear prediction would fail to match the reappearance of the occluded dot (Figure 1K). By what neural circuits can a human visual system produce a prediction for the case where the middle dot belongs to the group (Figure 1L)? This case is especially difficult because the group-based prediction for the middle dot (that its motion should be vertical) does not equal the motion of the group itself (diagonal).

Value of grouping: Improved predictions

Figure 1 suggests that the predictive performance of visual motion circuits is improved if the circuits incorporate information about grouping. This conclusion is based on the assumption that visual elements (like the dots) that behave coherently (e.g., move with the same velocity) will most often continue to behave coherently (at least over the short term). This Gestalt "common fate" assumption is reasonable in normal visual environments.

The transfer-of-binding problem

Even if two visual elements have common motion, they might belong to different objects. For example, suppose that two dots first move in different directions and later move with a common direction and speed. In that case, there is ecologically less reason to believe that they will continue having a common motion (which might just be coincidental) than for two dots whose motion had never been observed to differ. A visual system must be able to represent both cases: in the absence of contrary information, common motion should enhance grouping; but evidence of prior non-common motion should weaken grouping. This implies that visual systems should base their groupings on history as well as common motion. It also implies that a given dot has the potential to belong to many different groups (one at a time), consistent with its motion. Thus, a dot's actual belongingness, or bindings, must be represented *dynamically*, in a way that can be altered according to the dictates of the scene and the motion history.

How can a visual system use past bindings to predict and influence future bindings? This is the *transfer of bindings* problem. It may arise in many aspects of perception *(Kruschke & Fragassi, 1996)* and cognition *(Kahneman & Treisman, 1992)*. It is exemplified by the three-dot examples of Figure 1, but it is an inherent problem that arises because neurons and synapses cannot physically move to keep up with their moving or transforming inputs. Our model neural circuit for dynamic binding correctly performs transfer of binding for moving objects.

Methods: Neural circuit simulation

The model comprises five neural processing layers with a highly specific pattern of excitatory and inhibitory connection pathways (symbolized in Figure 2). The pattern of connections can be established by simple activity-dependent learning rules, in response to ordinary moving visual stimuli during a developmental period *(Hubbard & Marshall, 1996; Marshall, 1990, 1992, 1995; Schmitt & Marshall, 1995)*. Binding is effected by a model inhibitory interneuron gating circuit that causes feedforward excitatory afferent axonal pathways emanating from a common source neuron to compete for the "right" to carry signals, based on feedback signal strength. When a neuron e, representing a visual element, is bound to a neuron g, representing a group (or object), the $e \rightarrow g$ and $g \rightarrow e$ pathways are open, and other pathways $e \rightarrow g'$ and $g' \rightarrow e$ are gated shut by competition. This allows the spike pattern of neuron e to become synchronized to the spike pattern of neuron g.

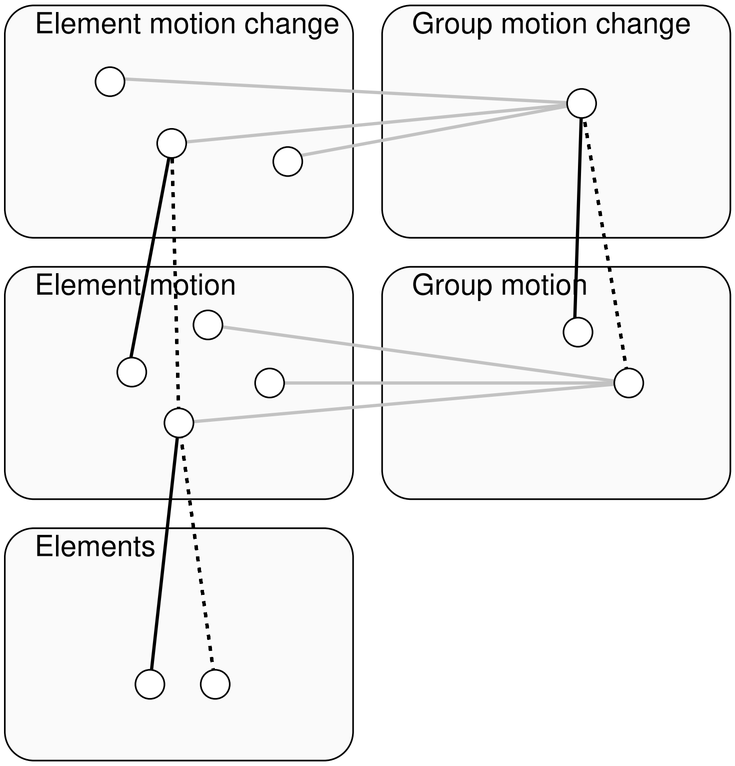


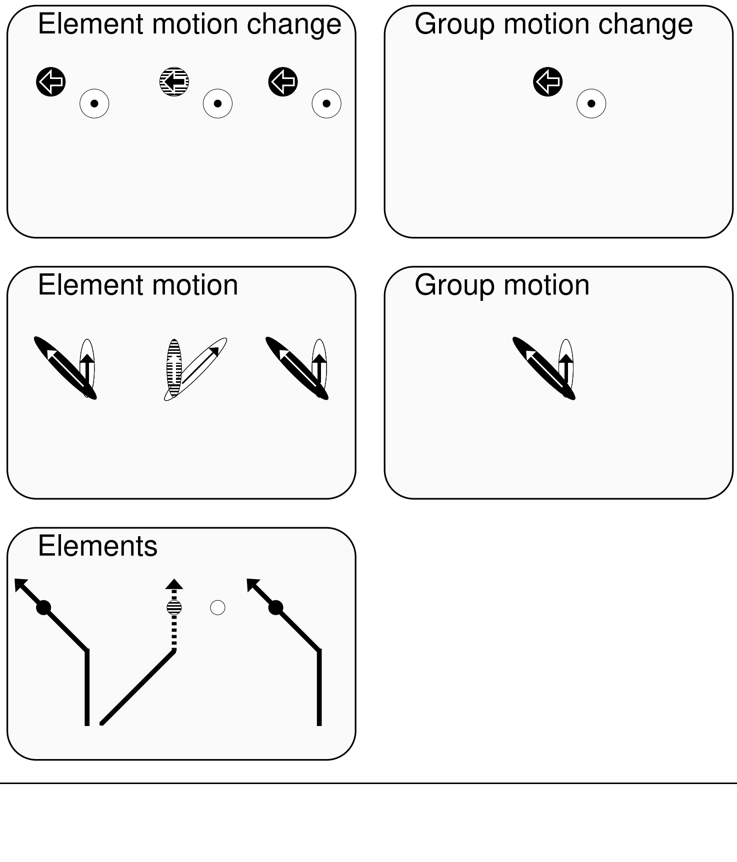
Figure 2: Neural architecture for relative motion perception, with transfer-of-binding. Five processing areas are shown. Input arrives in the Elements area. The Element Motion area contains simple bilocal motion-sensitive neurons, using time-delay pathways (dotted lines). (Other architectures for motion-sensitive neurons would work as well.) The Element Motion neurons can be dynamically bound to neurons in the Group Motion area, via binding pathways (gray lines). Second-order changes (e.g., curves or accelerations/decelerations) in the motion trajectories across the Element Motion and Group Motion areas activate neurons in the Element Motion Change and Group Motion Change areas. When a group motion change occurs, predictions for the new motion of each element are transmitted via reciprocal feedback pathways to the Element Motion Change neurons, further back to the Element Motion neurons, and further back to the Element neurons. Interneuronal binding circuits (not shown) implement competitive decision-making about temporal correspondence matches and motion assignments. Neuron spiking patterns (not shown) are transmitted between the areas and influence the decisions about which Group Motion neurons become active, so that elements that were grouped together at one motion step are more likely to remain grouped together at the next motion step.

Transfer of binding is effected by a mutual facilitation arising from synchrony of input spike firing patterns. When several visual elements are identified as belonging together to a group (or object), the neurons representing those elements become dynamically bound to a neuron representing the group; this "group" neuron then provides synchronizing feedback to the "element" neurons. Subsequently, after additional motion, the synchrony of the spike patterns from the element neurons enhances the likelihood that they will be represented together by a common group neuron at their new spatial positions.

Results: Predictive tracking based on group coherence

Figure 3 shows the results of a computational simulation of the model's behavior in response to the visual relative motion stimulus of Figure 1I. The simulation illustrates how the model uses information about the behavior of some elements of a group to predict the behavior of other elements of the group. The simulation's state is shown at the moment that the motion of the two outer dots has changed direction. The simulation generates a feedback signal that predicts the middle dot's location, based on *group* motion, rather than on element motion of the middle dot alone. The simulation, which used 2341 simulated neurons, was run in approximately 20 minutes on a Hewlett-Packard J210 workstation computer.

Figure 3: Simulation results: Time=3. The state of the simulated network at time=3 is shown. From time=1 to time=2, the two outer dots moved vertically upward, and the middle dot moved diagonally upward (heavy arrows in Element area). At time=3, the two outer dots (black dots in Element area) move diagonally upward, and the middle dot disappears, as if behind an occluder. Delayed excitation begins to activate two Element neurons at candidate locations for the occluded middle dot (two circles near center of Element area). Two Element Motion neurons are activated, and they bind to a common Group Motion neuron (black ellipses). The directional change of the group (from vertical to diagonal) activates a Group Motion Change neuron (black circle with leftward arrow). This sends feedback to the Element Motion Change neurons, which combines with candidate feedforward excitations to fully activate the two outer Element Motion Change neurons and to partially activate the middle Element Motion Change neuron (hatched circle). In turn, this partial activation feeds back excitation to the middle Element Motion neuron (hatched ellipse) and to the middle Element neuron (hatched circle). This feedback represents the network's prediction (hatched circle) that the motion of the (invisible) middle dot changes in accordance with the group motion, not as a mere ballistic extrapolation (open circle) of the middle dot's previous motion. If the firing pattern of the neurons representing the middle dot had not been synchronized with the firing pattern of the neurons representing the other two dots, then the prediction would *not* be influenced by the motion change of the two outer dots, and the prediction *would* be a linear extrapolation of the middle dot's prior motion.



Significance

This study has identified a computational problem, *transfer*-of-binding, that arises in visual relative motion perception, as well as in other areas. The distinction between binding and transfer-of-binding helps clarify the demands on biological visual systems in interpreting relative motion displays.

The model suggests a plausible neural mechanism for binding and a plausible functional role for spike synchrony in transfer-of-binding. These aspects of the model are neurophysiologically plausible for several reasons:

- the competitive interneuron circuit for binding uses only well-known and widely-accepted components;
- a consideration of specific spike patterns at individual neurons is not required; only synchrony *between* spike patterns is examined; thus, only a weak model of spiking is needed;
- spike synchrony provides only an amplification factor to the strength of neural signals; the activation of neurons depends only partially on the synchrony; thus, again, only a weak model of spiking is needed.

Our research has yielded several insights into the principles that determine how transfer-of-binding works and how visual relative motion perception works:

- Highly specific grouping/binding circuits are needed for precise relative motion perception.
- A new conception of visual binding, combining both gating *and* spiking synchrony, has been formulated.
- The problem of transfer-of-bindings, which has been an obstacle for prior models of motion perception, has been revealed and articulated.
- The first neural circuit that solves the transfer-of-bindings problem has been designed.
- Such a neural circuit can have a simple, regular, hierarchical structure.
- The neural circuit can establish, break, and re-establish the bindings for visual elements whose motion groupings change.
- The neural circuit can use the motion of the visible parts of an object to *steer* the representation of the motion of the invisible parts of the object.
- The hierarchical organization of the neural circuit allows the motion of an element to be represented in the reference frame of the motion of the group to which it belongs.

Acknowledgements

Supported in part by ONR grants N00014-93-1-0208 and N00014-93-1-1296. We thank Vinay S. Gupta and Viswanath Srikanth for useful discussions.

References

Hubbard RS, Marshall JA (1994) Self-organizing neural network model of the visual inertia phenomenon in motion perception. Technical Report 94–001, Department of Computer Science, University of North Carolina at Chapel Hill, 26 pp.

Johansson G (1950) *Configurations in Event Perception*. Uppsala: Almquist and Wiksells Boktryckeri AB.

Johansson G (1985) About visual event perception. *Persistence and Change: Proceedings of the First International Conference on Event Perception* Warren WH, Shaw RE (Eds.). Hillsdale, NJ: LEA, 29–54.

Kahneman D, Treisman A (1992) The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24:175–219.

Koffka K (1935) *Principles of Gestalt Psychology*. New York: Harcourt, Brace and Company.

Kruschke JK, Fragassi MM (1996) The perception of causality: Feature binding in interacting objects. *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates, 441–446.

Marshall JA (1990) Self-organizing neural networks for perception of visual motion. *Neural Networks* 3:45–74.

Marshall JA (1992) Unsupervised learning of contextual constraints in neural networks for simultaneous visual processing of multiple objects. *Neural and Stochastic Methods in Image and Signal Processing*, Chen SS, Ed. Proceedings of the SPIE 1766, San Diego, CA, 84–93.

Marshall JA (1995) Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity, and scale. *Neural Networks* 8:335–362.

Marshall JA, Alley RK, Hubbard RS (1996) Learning to predict visibility and invisibility from occlusion events." *Advances in Neural Information Processing Systems* 8, Touretzky DS, Mozer MC, Hasselmo ME (Eds.). San Mateo, CA: Morgan Kaufmann Publishers, 816–822.

Ramachandran VS, Inada V, Kiama G (1986) Perception of illusory occlusion in apparent motion. *Vision Research* 26:1741–1749.

Schmitt CP, Marshall JA (1995) Development of a neural circuit for motion grouping and disambiguation. Conference abstract, *Investigative Ophthalmology & Visual Science* 36(4):377.

Ternus J (1926) Experimentelle Untersuchungen über phänomenale Identität. *Psych. Forsch.* 7:81–136.