

MS number 1975

# Visual Motion Detection

**B K Dellen**, University of Goettingen, Goettingen, Germany

**R Wessel**, Washington University, St Louis, MO, USA

## Contact information:

Babette Dellen, Bernstein Center for Computational Neuroscience, University of Goettingen, Germany.

Tel: +49-551-517-6509, [bkdellen@bccn-goettingen.de](mailto:bkdellen@bccn-goettingen.de)

Ralf Wessel, Dept. of Physics, Washington University, St. Louis, MO 63130.

Tel.: 314-935-7976, Fax: 314-935-6219, [rw@physics.wustl.edu](mailto:rw@physics.wustl.edu)

**Key words:** motion, apparent motion, context, depth perception, direction, image segmentation, looming, motion contrast, motion repulsion, optic flow, retina, tectum, transparent motion, visual cortex.

Visual motion is essential for many behaviorally important functions such as camouflage breaking, detecting looming objects, and depth perception. Camouflage is a very effective survival tool, in which an animal adapts the color and texture patterns on its skin to match those of its habitat, thereby reducing the risk of being detected by preys or predators. This tactic, however, only works as long as the animal stays still. Once it moves relative to the background that it is hiding in, other animals can use their visual motion detecting apparatus to break the camouflage. Looming visual patterns are produced by a fast approaching object, and must therefore be detected reliably and well in advance of the expected collision in order to provide the animal with enough time to execute its escape behavior. Translational self-motion perpendicular to the line of sight produces apparent motion of stationary objects on the retina. The magnitude of this motion is related to the depth of the object. Therefore, estimating the apparent motion of different parts of the visual image can aid the brain in constructing the three-dimensional structure of the world. Such knowledge is important, for example, for navigational purposes.

In summary, having access to motion information is instrumental to an animal's chances of survival. However, motion information is not explicitly encoded at the level of the retinal input. Rather, motion has to be computed from the spatiotemporal intensity changes in the retinal image as sensed by the two-dimensional array of photoreceptors. Given the importance of visual motion for

biological function, it is perhaps not surprising that all animals with vision have mechanisms for motion processing.

### **Local motion processing**

There are basically two local strategies in which a moving intensity pattern can be distinguished from both static intensity patterns and random spatiotemporal intensity fluctuations. (i) The intensity at a particular location and at a given instant of time is compared with the intensity value at a neighboring location measured some time later (**Figure 1**). (ii) The intensity change in time at one location is divided by the intensity change in space at the same location. In the literature the two algorithmic strategies are referred to as the (i) correlation- and (ii) gradient-detectors. Multiple versions of the correlation detector have become popular in the neurobiological literature. The gradient detector has found more acceptances in the computer vision literature.

Objects in our visual environment are normally distinguishable by a number of cues such as luminance, shape, color, or texture, yet the percept of motion is qualitatively independent of the different cues. This cue independence of motion perception is paralleled by a similar cue independence of the responses of many motion-sensitive neurons in primate visual cortex and in avian tectum. Cue independence of motion perception may be a useful operational principle. To perform optimally in a variable environment, a motion analyzer should generalize across stimulus cues, thus encoding the motion of a stimulus regardless of the

cue that enables it to be seen. In avian tectum cue independence of motion computation is mediated by phasic signal transfer at the retinotectal synapse of tectal wide-field neurons. For this wide-field neuron the distinction between static stationary stimuli and dynamic spatiotemporal stimuli, such as motion, is independent of the details of the stimulus.

Our understanding of the cellular mechanisms of direction selectivity is most advanced in the vertebrate retina, where directional selective retinal ganglion cells were first discovered some 40 years ago. We now know that a myriad of mechanisms, including dendritic and synaptic dynamics and structure, together provide the neural substrate to potentiate the responses of retinal ganglion cells to motion in certain directions.

### **Global motion processing**

A comprehensive understanding of visual processing of motion requires one to go beyond local processing of information and to consider the important effect of context. A neuron responds to visual stimuli presented within an area in the visual field, which traditionally is referred to as the “classical” receptive field. The neuron does not respond to visual stimuli presented outside this area, which is referred to as the “surround”. Importantly, however, the response to stimuli presented in the classical receptive field may change when additional stimuli are presented in the surround, i.e., the response depends on the stimulus context. Efforts to clarify the psychophysics and physiology of contextual effects in visual

motion analysis are collectively referred to as studies of global motion processing.

An important current area of research in global motion processing is transparent motion analysis (**Figure 2**). Motion transparency is a situation where two overlapping surfaces move past each other, and as a result the same - or very close - retinal locations may contain two different local motion parameters. In order to estimate the motion properties of the two surfaces, local motion signals have to be first segmented, i.e. assigned to either surface. Due to the spatial overlap of the surfaces, however, this segmentation process is nontrivial. While in many natural settings non-motion cues such as contrast, color, texture, etc can aid the brain with the segmentation task, there are important scenarios such as camouflage, in which the brain has to rely solely on motion information to partition the visual image into regions representing independently moving surfaces.

In laboratories, such conditions can be simulated using transparent random-dot patterns, where small dots distributed randomly across the visual space move in one of two possible directions, thereby constituting two groups. The location of a dot, however, does not predict its group identity. Further, the brain only has a noisy version of each direction of motion available. Thus statistical methods have to be employed for this unsupervised classification problem. For instance, accurate estimation of the background motion from an image velocity field can be

accomplished through an iterative cooperation between two modules: one that specializes in calculating a weighted average velocity, and another one calculating a velocity contrast map. Interestingly, this approach finds an analogy in the vertebrate tectum-pretectum system and suggests that the optic tectum is in charge of calculating a motion contrast map, while the pretectum computes background motion, i.e., the weighted average velocity.

“Motion repulsion” is a visual illusion that occurs during the perception of transparent motion. Motion repulsion is the perceived enlargement of the angle between the directions of motion of two transparently moving patterns. It has been hypothesized that this illusion is mediated by lateral inhibition between directionally selective units. An alternative explanation, however, relies on the statistical properties of the motion transparency problem alone. Formulating the problem as an unsupervised clustering problem reproduces motion repulsion.

Another important example of global motion processing by neurons is their sensitivity to motion contrast or relative motion between different parts of the visual image (**Figure 3**). For example, some cells in the deep layers of the avian optic tectum show suppression of response to a moving stimulus in their excitatory receptive field when it is accompanied by a second stimulus that is moving in the same direction, as long as the latter is within the inhibitory surround. Some neurons in the middle temporal cortex of primates show similar

properties. The underlying computation requires spatial comparison of visual signals, presumably mediated by lateral and/or feedback connections.

An algorithm of motion-contrast computation without motion sensors has been proposed. The algorithm classifies the visual scene into coherent and incoherent motion and determines the relative velocity of the moving parts of the visual scene. Remarkably, the algorithm predicts the measured response to motion contrast as a function of the relative angle of object motion from avian tectum and primate area MT.

### **Neural structure and motion processing**

It has long been hypothesized that structure and function are correlated in neural systems. The avian retino-tecto-rotundal pathway plays a central role in motion analysis and features complex connectivity. While the retinal ganglion cells report local image contrast, the tectal wide-field neurons fire more vigorously when facing small moving stimuli. Postsynaptic neurons in the thalamic nucleus rotundus are selective for the direction of motion of an object as well as looming patterns.

These differences in representations of visual stimuli are paralleled by qualitative structural differences of the neuronal elements in the three stages (**Figure 4**). A dense array of retinal ganglion cell axon terminals form a topographic map and synapse onto dendritic endings of tectal wide-field neurons. In contrast, tectal

wide-field neurons have large circular dendritic fields with sparsely distributed dendritic endings. Interestingly, the point-to-point topography of the retino-tectal projection is replaced by a more abstract topography in the tecto-rotundal projection that appears to be “interdigitating”, i.e., a specific locus in the nucleus rotundus receives input from a sparse population of tectal wide-field neurons throughout the entire tectum.

A computational investigation revealed that the tectal wide-field neuron’s structural properties enhance the neuron’s sensitivity to change, a behaviorally important stimulus attribute, while preserving information about the stimulus location in the tectal wide-field neuron population activity. The computational investigation showed further that the subsequent interdigitating tectal projection to rotundal neurons allows the decoding of the direction of motion from the sparsely sampled population of tectal wide-field neurons in a two-stage process. However, the study could not address why the tecto-rotundal projection evolved towards sparse sampling, as opposed to a one-stage projection with dense sampling, which produces similar computational properties. Interestingly, however, a wiring efficiency argument provides a plausible constraint to bias the evolution towards the sparse tecto-rotundal projection approach for motion estimation.

It is tempting to speculate that a computational rationale for sparse connectivity similar to the one hypothesized for the retino-tecto-rotundal pathway will be found



for other prominent computational pathways, such as the mammalian cerebral cortex, where only a small fraction of all possible connections between neurons physically exist, even within a local area.

### **Motion processing in visual cortex**

The geniculo-cortical pathway of the mammalian brain plays an important role in the processing and extraction of visual motion. The majority of retinal ganglion cells ascend via the dorsal part of the two lateral geniculate nuclei (LGN) of the thalamus to the primary visual cortex (V1). Then, the processing stream continues to higher visual areas, including the middle temporal area (area MT). Many neurons in V1 and MT are sensitive to motion and lesions of these areas severely reduce the ability of the animal to detect motion. These observations suggest that the geniculo-cortical pathway is highly involved in visual motion analysis.

Traditionally, neuroscientists have attempted to classify V1 neurons into two types (**Figure 5**). Within this classification scheme, one type, the so-called “simple cell”, possesses periodically arranged and oriented excitatory and inhibitory areas, which makes them sensitive to the spatial phase, spatial frequency, and orientation of a drifting grating. The other type, the so-called “complex cell”, is largely insensitive to the spatial phase of a drifting grating. To date, much research has been based on the working hypothesis that visual cortical neurons are hierarchically organized with simple cells feeding into

complex cells. It is important to note however that the experimental evidence for or against distinct simple- and complex-cell classes and its hierarchical arrangement has been numerous but not conclusive.

The product of an oriented sine function with a Gaussian envelope provides a useful model to reproduce the directionally-selective receptive field of simple cells. In the next step, model simple cell responses are nonlinearly combined to give complex-cell responses. These responses represent the projections of the stimulus velocity vector on different axes and need to be combined to obtain meaningful estimates of the full stimulus velocity. The latter computation may take place in area MT. Neurons in area MT show a strong directional selectivity and are possibly involved in the processing of global motion patterns and figure/ground segregation.

Interestingly, a mathematical analysis showed that local relative-motion computation is inherent in the nonlinear transformation from simple- to complex-cell responses. Simulations that quantitatively reproduce measured complex-cell responses in both cat and monkey to a variety of relative-motion stimuli support the mathematical inference. Importantly, this computation of relative motion does not require an intermediate neural representation of local velocities and does not require lateral or feedback interactions within a network.

The hierarchical model of visual cortical motion processing is in good agreement with a large body of experimental data. However, it is a local model. Velocities are extracted from a small area of the visual field only. For the simultaneous presentation of multiple stimuli, the hierarchical model loses most of its predictability. Already in V1, neurons exhibit sensitivity to orientation and motion contrast between a stimulus in the classical receptive field and the surround, potentially contributing to perceptual pop out and image segmentation. Furthermore, recent studies suggest that cortical processing is less local than previously thought. These properties of the visual cortex might be attributed to the abundant lateral and feedback connections, which innervate the cortical layers. For instance, less than 5% of excitatory synapses on V1 neurons originate from the LGN, while over 95% of the excitatory synapses, even in the geniculate-recipient layer in area V1, come from other cortical neurons and other nuclei. For instance, feedback from higher visual areas may represent a prime candidate to mediate contextual effects.

Contextual influences from outside the classical receptive field may fundamentally change our view of visual cortical motion processing. The challenge ahead is to design global theories of the brain, in which locality is compatible with global processing strategies.

## **Further Reading**

- Albright TD, Stoner GR (2002) Contextual influences on visual processing. *Ann Rev Neurosci* 25: 339-379.
- Andersen RA (1997) Neural mechanisms of visual motion perception in primates. *Neuron* 18: 865-872.
- Borst A, Egelhaaf M (1989) Principles of visual motion detection. *TINS* 12: 297-306.
- Clifford CWG, Ibbotson MR (2003) Fundamental mechanisms of visual motion detection: models, cells and function. *Progress in Neurobiology* 68: 409-437.
- Dellen BK, Clark JW, Wessel R (2004) Motion-contrast computation without directionally sensitive motion sensors. *Phys Rev E* 70:031907.
- Dellen BK, Clark JW, Wessel R (2005) Computing relative motion with complex cells. *Visual Neuroscience* 22: 225-236.
- Egelhaaf M, Kern R, Krapp HG, Kretzberg J, Kurtz R, Warzecha AK (2002) Neural encoding of behaviorally relevant visual-motion information in the fly. *TINS* 25: 96-102.
- Giese MA, Poggio T (2003) Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience* 4: 179-192.
- Hildreth EC, Koch C (1987) The analysis of visual motion: From computational theory to neuronal mechanisms. *Ann Rev Neurosci* 10: 477-533.

- Luksch H, Khanbabaie R, Wessel R (2004) Synaptic dynamics mediate sensitivity to motion independent of stimulus details. *Nature Neuroscience* 7: 380-388.
- Mahani AS, Carlsson AE, Wessel R (2005) Motion repulsion arises from stimulus statistics when analyzed with a clustering algorithm. *Biol Cybernetics* 92: 288-291.
- Mahani AS, Khanbabaie R, Luksch H, Wessel R (2006) Sparse spatial sampling for the computation of motion in multiple stages. *Biol Cybernetics* 94: 276-287.
- Mahani AS, Wessel R (2006) Iterative cooperation between parallel pathways for object and whole-field motion. *Biol Cybernetics* 95: 393-400.
- Nakayama K (1985) Biological image motion processing: a review. *Vision Res* 25: 625-660.
- Palmer SE (1999) *Vision science*. MIT Press.
- Sereno ME (1993) *Neural computation of pattern motion*. MIT Press.
- Smith AT, Snowden RJ (eds) (1994) *Visual detection of motion*. Academic Press.
- Taylor WR, Vaney DI (2003) New directions in retinal research. *TINS* 26: 379-385.
- Wandell BA (1995) *Foundations of Vision*. Sinauer.
- Watanabe T (ed) (1998) *High-level motion processing*. MIT Press.
- Zanker JM, Zeil J (eds) (2001) *Motion vision: Computational, neural, and ecological constraints*. Springer.

## Figure Captions

**Figure 1** Schematic representation of the correlation-based motion detection. The snapshot of an intensity edge moving with a constant velocity in x-direction is shown. The intensity pattern within a small window of the image is compared with the intensity pattern within another window at a neighboring location some time  $\tau$  later. If the intensity patterns are sufficiently similar, or ideally identical, the measured velocity is equal the distance  $\delta x$  between the two windows divided by the time delay  $\tau$ . The similarity of the intensity pattern may be evaluated via the correlation coefficient.

**Figure 2** An example of transparent motion. Two overlapping transparent random-dot pattern are moving on top of each other in different directions. Hence, the location of a dot does not predict its group identity, and a single point in space might contain two different velocities.

**Figure 3** An example of relative motion or motion contrast. An opaque random-dot square is moving to the right in front of random-dot background, which is moving to the left. Thus, the random-dot square is only defined by its motion relative to the background.

**Figure 4** A schematic of the avian retino-tecto-rotundal connectivity pattern. Retinal ganglion cell (RGC) axons project to the dendritic endings of tectal (OT) wide-field neurons. The tectal wide-field neurons in turn project with an interdigitating topography to the nucleus rotundus (Rt).

**Figure 5** Schematic of the receptive-field structure of a cortical simple and complex cell. Each simple cell filters the visual input by means of an oriented spatiotemporal sine function multiplied with a Gaussian envelope. Then the output of the filters is rectified to ensure positive-valued responses. The responses of two simple cells whose sine functions are phase shifted by  $\pi/2$  are first squared and then added to create the complex-cell response.

Figure 1

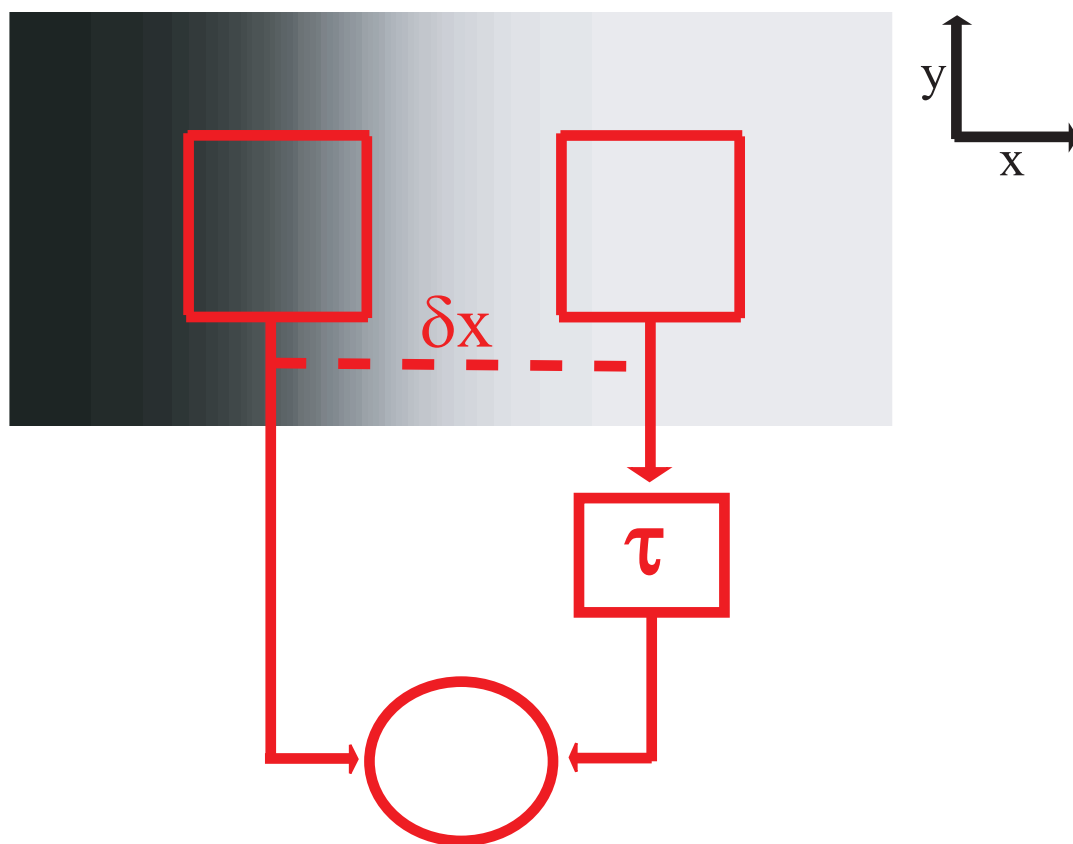




Figure 2

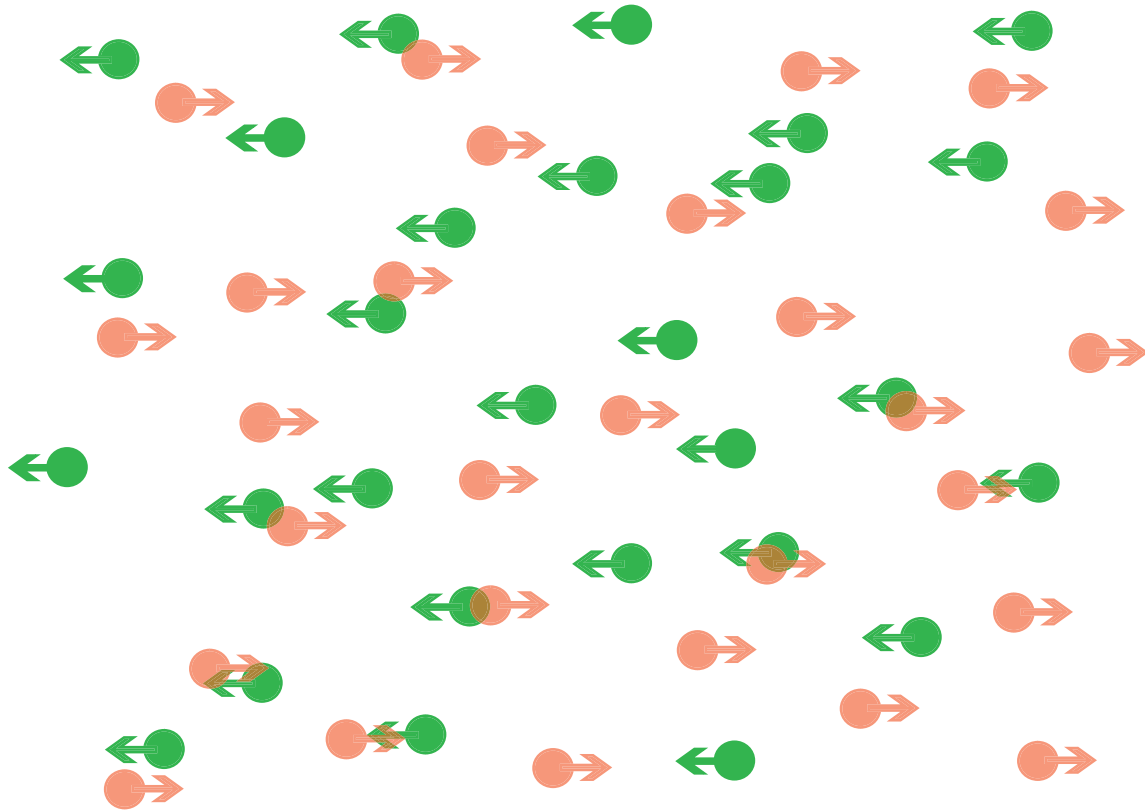


Figure 3



Figure 4

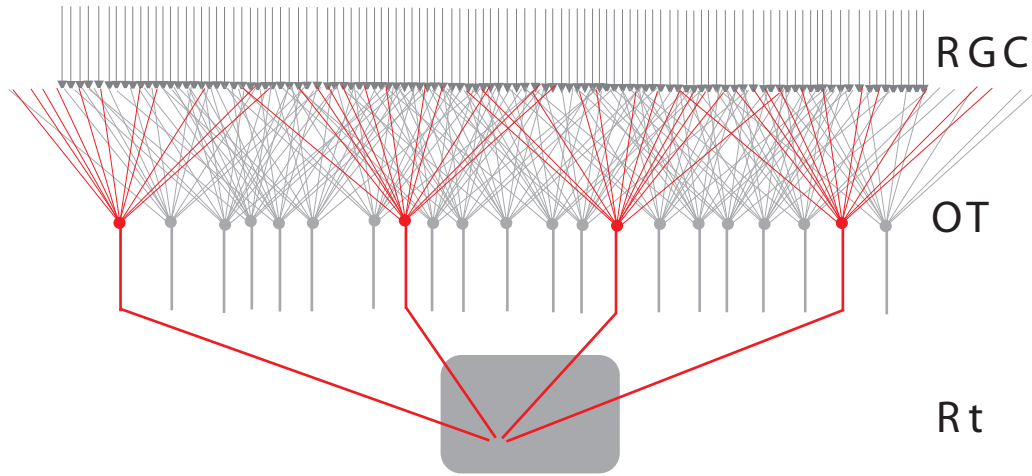


Figure 5

