UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

UNIVERSITY OF PADOVA
DEPARTMENT OF INFORMATION ENGINEERING

CORSO DI LAUREA SPECIALISTICA IN BIOINGEGNERIA

TESI DI LAUREA

A NOVEL METHOD FOR COMPUTING MOTION DISCONTINUITY

Relatore: Prof. Alfredo Ruggeri

Correlatrice: Prof.ssa Lucia M. Vaina

Brain and Vision Research Laboratory
Department of Biomedical Engineering
Boston University (USA)

Laureando
Davide Adamoli

Padova, 13 Aprile 2010
Anno Accademico 2009/2010

A Francesca

Contents

Sommario (in Italiano)

Le discontinuità del movimento sono definite dalla presenza di regioni adiacenti con moto differente per verso e/o velocità. L'individuazione di discontinuità del movimento è considerata dipendente dall'integrazione di informazione sul moto su scala non-locale.
Esperimenti basati di cinematogrammi dinamici a punti casuali sparsi hanno mostrato che in uno schermo rumoroso con un piccolo numero di punti che si muovono coerenti, i soggetti testati hanno la percezione di un movimento globale verso una direzione. I punti-rumore vengono ignorati e non contribuiscono al moto puro. In altri esperimenti dove nello stesso schermo era presente una discontinuità del movimento, i soggetti “vedevano” un'invisibile linea di discontinuità.

Allo scopo di verificare l'ipotesi che l'integrazione del moto e il calcolo delle discontinuità fossero legati, pazienti post-lesione sono stati testati con uno stimolo che poteva essere utilizzato per testare sia la percezione della coerenza che il rilevamento delle discontinuità, trovando una doppia dissociazione in soggetti che riuscivano in un test ma risultavano deficitari nell'altro. Questi risultati suggerirono che i processi non fossero legati, così contraddicendo molti modelli.

In questa tesi si analizza il problema del rilevamento delle discontinuità del moto nei suoi aspetti formali, considerando la letteratura esistente, e viene proposto un nuovo algoritmo biologicamente realizzabile basato su di un modello originale di Nakayama e Loomis (1974). Il modello, fondato su di un meccanismo di centro-contorno, utilizza solo la componente normale del flusso ottico retinico, che si sa essere disponibile inizialmente nel sistema visivo e calcola un “valore di convessità” scalare per ciascun luogo dell'apertura visiva. La funzione scalare sviluppata ha alti valori alle discontinuità e bassi valori altrove, cosicché, mettendo a soglia i valori, le discontinuità possano essere isolate.

La funzione Convessità è stata implementata e testata in simulazioni che ricreavano le condizioni delle vere sessioni di test (il codice integrale è incluso nell'Appendice). I risultati sono riportati e discussi, confrontandoli con le normali prestazioni umane.

Summary (in English)

Motion discontinuities are characterized by the presence of adjacent areas with different motion, in direction and/or speed. The detection of motion discontinuities has been considered to be dependent on integration of motion information over a non-local region. Experiments with dynamic sparse random-dot kinematograms showed that in a noisy display with a small number of coherently moving dots tested subjects have the perception of a global motion towards one direction. Noise dots are ignored and do not contribute to the net motion. In others experiments where in the same display was present a motion discontinuity, subjects would “see” an invisible line of discontinuity.

In order to test the hypothesis that motion integration and discontinuity computation are coupled, post-lesion patients were tested with a stimulus that could be used to test both coherence perception and discontinuity detection, finding a double dissociation in subjects that could perform one test but where deficient in the other. These results suggested that the two perception processes are not coupled, thus contradicting several models.

In this thesis the problem of detecting motion discontinuities is analyzed in its formal aspects, considering the existing literature, and a new biologically feasible algorithm is proposed based on an original model of Nakayama and Loomis (1974). The model, based on a center-surround mechanism, uses only the normal component of the retinal optic flow, named normal flow, that is known to be available early in the visual system and it calculates a scalar “convexity value” for each location of the visual aperture. The scalar function developed has high values at discontinuities and low values elsewhere, thus by thresholding discontinuities can be isolated.

Convexity function was implemented and tested in simulations that recreated real test session conditions (full code included in Appendix). Results are reported and discussed, comparing them with normal human performances.

Chapter 1 **Introduction to the problem**

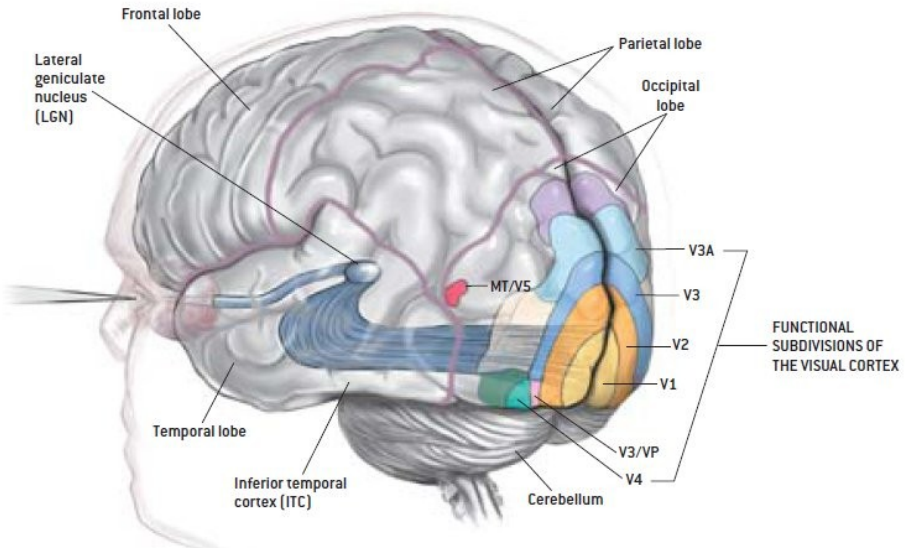
> "*Almost I don't see any difference in how things move. (…) I have an eerie feeling that I'll bump into them.*" - AMG, 53 years old cerebral stroke patient

> "*Non vedo quasi differenze in come gli oggetti si muovono. (…) Ho un sinistro presentimento che ci sbatterò contro.*" - AMG, paziente di 53 anni che ha subito un'emorragia cerebrale

Among the many functions of the **Human Visual System** that let's us perceive the surrounding environment and interact with it, two basic tasks that are performed at a very early stage are the computation of **motion coherence** and the detection of **motion discontinuities**. This thesis aim is to focus on the properties of these two functions and, specifically, to propose an explanation on how motion discontinuity detection could work in a biological implementation.
Previous work on this argument is analyzed and compared to data collected from tests to healthy volunteers and patients with syndromes, to propose a model of the motion discontinuity and a coherent schema of the detection process.

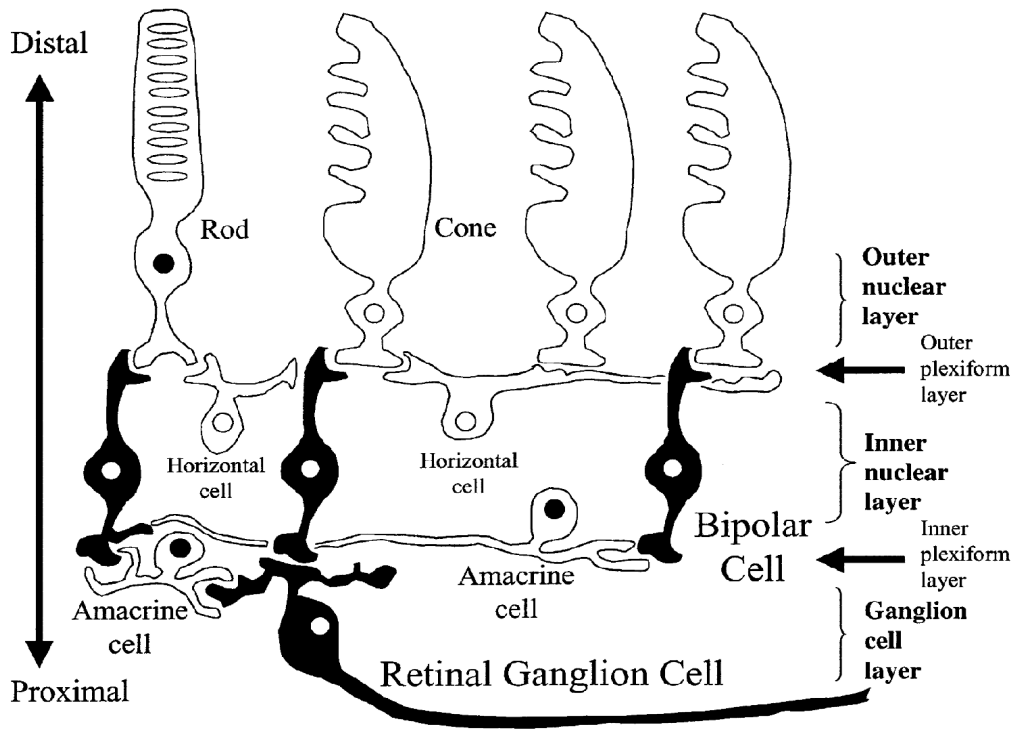
1.1 Human Visual System

As "Human Visual System" we intend the whole biological structure involved in the visual cognition process that begins in the retina and ends in the brain higher areas, just before the conscious cognition of the visual frame.



A representation of the human visual system from Logothetis (1999).

Light enters in the eye through cornea, pupil and lens to the retina, a complex nerve tissue composed of three layers: outer nuclear layer made of photoreceptors rods and cones; inner layer made of bipolar, horizontal and amacrine cells; ganglion cell layer.



Schema of retinal layers from Clifford, Ibbotson (2003)
Light comes from proximal and proceeds through cell structures to the light-receptive parts of rods and cones.

In each human eye there are approx. 120 million rods and 6 million cones, for 1.2 million fibers, thus the ratio between photoreceptors and ganglion cells is about 105:1 (Ganong, 2006).

The *fovea* has a diameter of 0,5 mm, is located in the center of the retina and has the highest visual acuity (25 arcseconds). In this particular region of the retina, that covers only 2° of the visual scene, there are almost only highly packed cones in a ratio 1:1 with ganglion cells, while in the retina periphery there could be 200 cones and rods for each ganglion cell. There are about 35.000 cones in the *fovea*.

Horizontal cells transmit signals between rods, cones and bipolar cells. Bipolar cells transmit signals from photoreceptors to both amacrine cells and ganglion cells. Amacrine cells make synapses with bipolar cells, ganglion cells or other amacrine cells.

There are three types of retinal ganglion cells:

- Type W: 40% of the total, wide receptive field, very reactive for directional movement, speed of impulses: 8 m/s;
- Type X: 55% of the total, narrow receptive field, speed of impulses: 14 m/s, they transmit the visual image;
- Type Y: 5% of the total, wide receptive field, high speed of impulses: 50 m/s, they react to rapid variations of the visual image (movements or brightness changes), although with no precision on where the variation occurred. (Guyton and Hall, 2002)

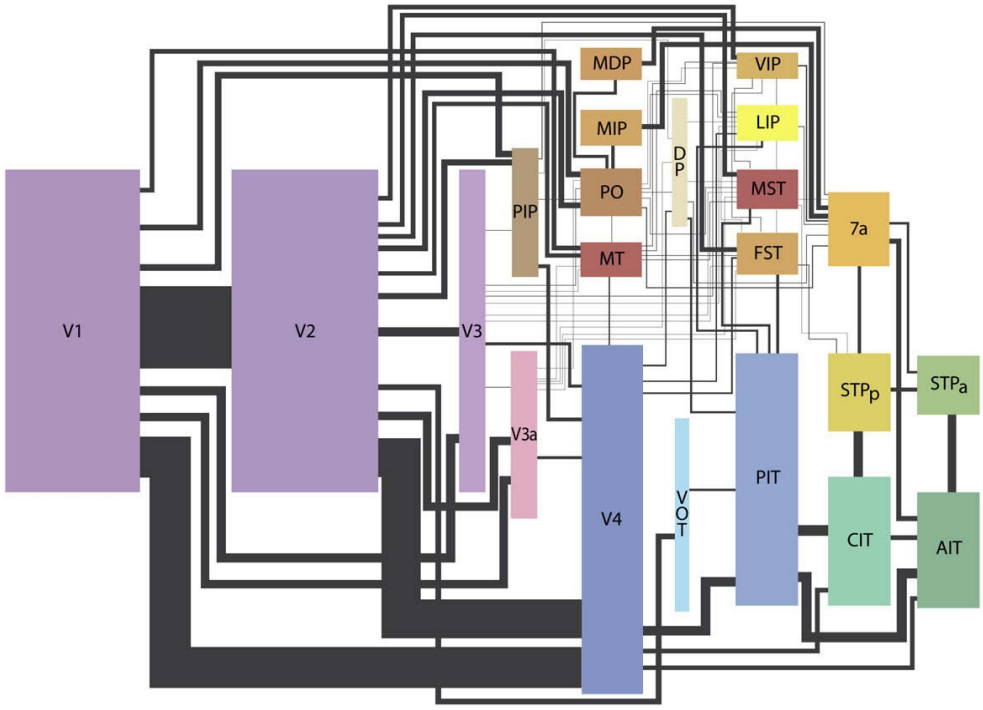
Studies on rabbit retina showed two types of direction-selective Retinal Ganglion Cells (DSRGC): cells that respond to movement along a preferred direction (directione-selective). Different DSRGC are selective to different directions. These types of DSRGC are:

- On-DSRGC, that reacts to the movement of bright edges;
- On-Off-DSRGC that reacts to the movement of both bright and dark edges in the image (Clifford and Ibbotson 2003).

Other studies on rabbit retina reported 4 subtypes of On-Off-DSRGC each one responding preferentially to motion in one of the four cardinal points (upwards, downwards, forwards and backwards) and that each point of the retina is covered by these four subtypes of DSRGC. Direction selectivity would be generated by the starbust amacrine cells (SBAC)

that provide both inhibitory and excitatory inputs to DSRGC to produce direction selectivity (Taylor and Vaney 2003).

Electrical signals are carried from retina to the visual cortex by the optic nerve to the lateral geniculate nuclei (LGN). Then visual information moves to V1 area (Primary Visual Cortex) and then distributed to several other regions as depicted in the following picture from Wallish and Movshon (2008).



A scaled representation of the Cortical Visual Areas of the macaque, taken from Wallish and Movshon (2008). Each area is proportional to its cortical surface, the thickness of the connections is proportional to the estimated numbers of fibers in the connection. Original version by John Maunsell in 1998.

The Primary Visual Cortex or V1 or striate cortex is located in the occipital lobe and is accounted for a number of linear and non-linear functions computed at a local level. Simple cells have receptive fields oriented in space and time and respond preferentially to a specific direction of movement. Complex cells have non-linear responses such as inhibition of responses in non-preferred directions, multiplicative and squaring operations in the preferred direction.

Middle Temporal area (MT or V5) and Medial Superior Temporal area (MST) are accredited for a great number of higher functions explained in section 3.10.

1.2 Visual motion perception and measurement

The final objective of the Human Visual System is to reconstruct the real 3D world from a 2D projection of light onto retina, so that the individual can interact with it.

The motion information on which reconstruction relies is inferred from the pattern of changing light intensity on the retina, a direct projection of the light signal coming from illuminated objects that is focused by the eye structure. The light signal is then transformed into an chemical/electrical signal and then is computed through various steps.

The first step of the process is to compute a **2D velocity field**, i.e. a field of velocity vectors (defined by direction and speed) assigned to the elements of the image, from the bare signal coming from the retina.

The second step is to organize this velocity field to extract distinct moving objects in the scene and to compute coherent movements.

Finally the higher task is to reconstruct the 3D world information so that it can be used in everyday life (that we will not address here).

We can measure motion perception ability of humans or primates through a series of lab tests, that are standard, replicable and normalized to a population of healthy volunteers as control. Motion discontinuity test is described in Chap. 4. Usually in these test the subject is given a monitor to watch and a possible yes/no or up/down or left/right possible response. The test proceed in a series of showings to the subject, where he/she has to give an answer on what he/she perceives. Gradually, properties of the test are changed for every repetition, such as signal/noise ratio, speed, dimensions, until the subject fails to give the correct answer. That point marks the subject perception level.

Other invasive tests, done on macaque or rhesus monkeys due to their similarity with humans, but even on other mammals (like cats), use surgery to set lesions at specific locations, so that the difference in motion perception pre/after lesion can be measured and theories proved o declined. Monkeys in particular can be trained to motion perception tests, thus giving an excellent feedback after lesion (see Newsome and Paré 1988).

Finally, lethal tests performed on anesthetized macaque monkeys, as for example Majaj-Carandini-Movshon (2007), aim to isolate single neurons, map their receptive fields, test them with a given input (gratings, plaids) and measure their evoked response. At the end of the experiments monkeys are killed.

1.3 Motion coherence

Motion coherence is a basic function implemented to perceive in the image flow any coherent motion along one or more directions. It is a typical integration task, since it adds up all motion directions perceived in the visual scene and detects predominant directions among the noise or static elements.

The direction of the motion field cannot be derived by looking at individual spots, but depends on the integration of motion information over a large area.

Such a function is essential to get early information from the scene, without conscious elaboration, to detect and coherent movement in the environment. We could imagine, for example, a prey in its daily feeding seeing in the grass some coherent movement. It could be a predator approaching, thus having a fast detection system would rise the chances of survival.

1.4 Importance in robotics and medicine

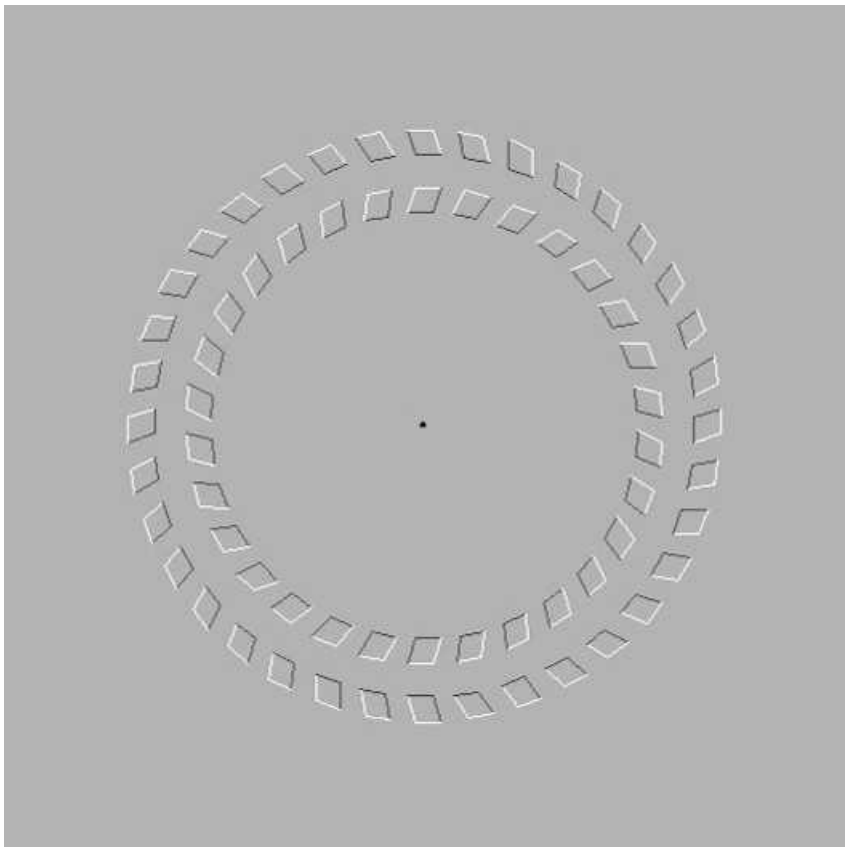
Understanding how a biological system works is particularly important to design efficient artificial vision systems. Though many different computations based on functions not computable in a biological environment are possible in digital imaging, usually biology shows the most efficient and parallelizable way, thus guiding towards new approaches.

In medicine much remain unclear about human brain and specifically about the visual system. Even the specific visual functions associated with cortical areas are not well defined. The purpose of this work is to clarify whether the possibility of a motion discontinuity detection like the one illustrated here is reasonable or not. If so, further can study more specifically the question with tests on patients, adding a piece of information to the great map of human brain.

1.5 Illusions

As final paragraph for this introduction, we report some famous illusion images that lever to Human Visual System deficits to induce false but fascinating perceptions in the viewer. At the end of this thesis the reader should be able to understand why this illusions occur.

The false movements



The reader fixes the dark dot in the middle of the two concentric circle-like structures, then draws the image near the eyes. He/She perceives two (false) movements of the circles. The external one seems to turn clockwise, the internal one anticlockwise. The mind cannot oppose to it since these apparent movements are computed prior to conscious elaboration, and are generated by a wrong reconstruction of the velocity field from the visual scene (normal flow).

Barber pole, ca. 1938., North Carolina Museum of History (www.wikipedia.org)

The barber pole illusion.

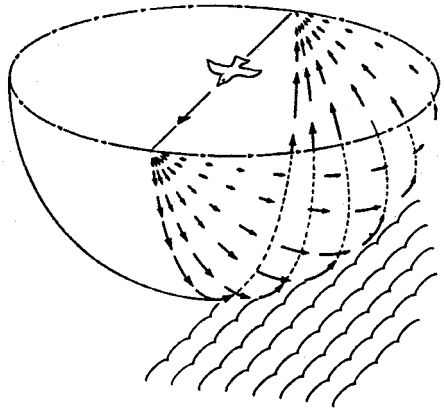
A spinning barber pole produces a motion perception illusion in which the stripes appear to be traveling down the length of the pole, rather than around it.

Again, the computed velocity field of least variation, computed on the basis of the normal velocity field, is not the true velocity field. If bars were drawn with different textures, instead of being uniformly colored, the true motion would be easy to discriminate.

Chapter 2 **The problem of detecting Motion Discontinuity**

2.1 Optical Flow

The human visual system perceives the world through a 2D projection of the 3D environment onto the retina. This projection is the spherical representation of the environment in the observer and is called **optical flow**. Each point on the sphere correspond to a unique environmental point. However, it is clear that the two-dimensional retinal image provides sufficient information for a 3D reconstruction, through monocular processing of optical velocities. It's to be noted that a uniform flow of the environment does not produce an uniform flow on the retina, being the flow faster in the central region and slower at borders.



Optical flow for a bird flying over the ground (Gibson, 1966)

The relation between velocity and displacement of object projection provides the data for the calculus of the object's depth.

We notice that **Velocity = Speed + Direction**.

This means that velocity vector is the sum of two different **pieces of information** which may not be available at all times. In some visual structures is computed speed, in others is computed direction. In mathematical terms: $\vec{v} = |v| \cdot \vec{u}$

Retina cells cannot simply convert the light signal into an exact electrical replica due to the aperture problem (see next paragraph), thus optical flow needs to be computed using only the visual information available through a global integration. Many different models were proposed to solve this problem, yet none was proved directly. Techniques to detect the

flow vectors from two subsequent image frames are usually:

- spatio-temporal derivatives
- correlation-based algorithms

Moreover, even with all information possible, optical flow computation would be uncertain near motion discontinuities, i.e. places where two different surfaces moving in two different directions overlap generating a discontinuity in the field of velocity vectors (*flow field*). This happens because spatial integration of local flow at motion boundaries positions leads to erroneous detection. Most visual models proposed that use global integration fail to provide explanation on how human visual system is capable of great precision in detecting and solving motion discontinuities in the flow field.

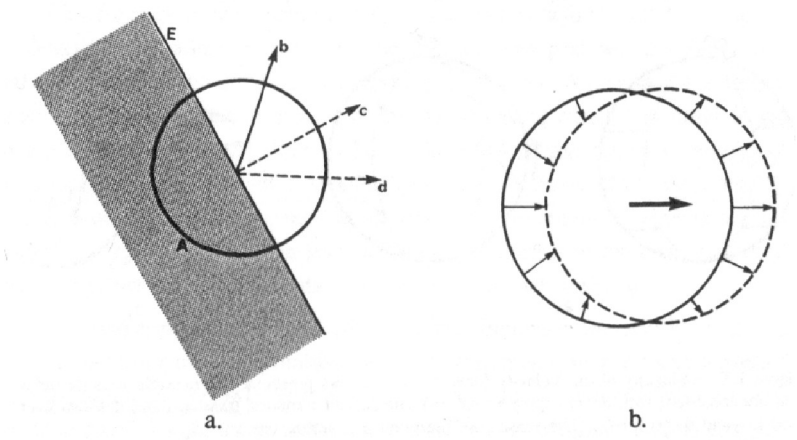
Finally, motion has to be distinguished in

- self motion: due to the movement of the observer;
- local motion: parts of the visual field that move independently.

In this work we will always consider self motion = 0 (observer is not moving).

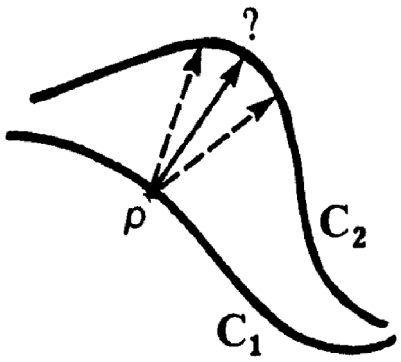
2.2 Aperture problem

Given a very small circular aperture of the visual scene, the aperture problem posits that only the component of optic flow normal to the local intensity gradient edge can be computed.



From Hildreth (1983) - Aperture problem
a. possible movements along the moving edge E through the local aperture A: only the perpendicular one can be computed;
b. the circle is translating to the right, but only local perpendicular components of velocity can be obtained from the changing image.

Since there's no difference between points of the line of the local intensity gradient, it's impossible to distinguish each point. Thus, given the original line and its evolution after a movement, the new point location it's uncertain. This leads to the fact that the point movement is ambiguous, the transverse component of velocity cannot be perceived.
So, at a local level we can only calculate normal velocity, i.e. the velocity vector normal to the local intensity gradient.



*From Hildreth (1983) - Curve C_1 evolves in curve C_2, but the velocity vector of the point **p** is ambiguous*

2.3 Normal Flow

The aperture problem indicates that in fact we don't have an *optical flow* available for our calculations, but a poor version of it that includes only the projection of velocity to the vector normal to the local visual edge. The component of velocity along the edge remains undetected. For every velocity vector in the optic flow we have its corresponding normal projection, the whole vector set is called **normal flow**. Normal flow is computed locally and does not require any global information.
It is possible to integrate normal flow with various methods and obtain a velocity field very similar to the original optic flow, but what human visual system really perceives is always just normal flow.
Local motion measurements are obtained from the changing image, Marr and Ullman (1981) proposed it happens at locations of significant intensity changes. Such intensity changes form a contour of the image and that contour would provide the local edge orientation for motion measurements. Thus, these motion measurements would provide the component of velocity perpendicular to that local edge.

Another location where measure motion are the so called zero-crossings. Marr and Hildreth (1980) proposed an operator for the initial filtering of the image, a Laplacian of a Gaussian, $\nabla^2 G$, approximated in shape by the difference of two Gaussian functions. They suggested that in the primate visual system, the convolution of the retinal image with $\nabla^2 G$ is represented in the output of the class of retinal ganglion cells referred to as the X-Cells. Simple cells in primate cortex have receptive fields that respond to movements of edges in preferred directions, thus Marr and Hildreth proposed that a class of simple cells may assume a role in the detection of segments of the zero-crossing contours. Later, Poggio (1983) proposed a second model based on the hypothesis of a class of simple cells that detect moving zero-crossing segments.

2.4 Motion discontinuities

Motion discontinuities are changes in the local flow field. If we have a surface moving in a visual scene, we'll have a corresponding field of coherent velocity vectors in the velocity flow field. When two surfaces with different motions overlap, they generate a motion discontinuity since we'll have on one side coherent velocity vectors with a specific direction and on the other side velocity vectors with another direction.

So, motion discontinuities imply a **local** computation of some visual quantity such as motion direction or speed.

Motion discontinuities usually define motion boundaries that are related to object's boundaries in the visual scene. A moving object can be easily and precisely perceived even if it's texture is mimetic with background.

There are many visual tests that confirm the capacity of the human visual system to extract object boundaries starting from motion information alone. This leads to the capacity of the visual system to extract motion discontinuities given a visual flow field, that is (for motivations above described) extract motion discontinuities from simple normal flow.

2.5 Detecting motion discontinuities

There are basically 3 possibilities for detecting motion discontinuities:

a) detecting discontinuities prior to computation of the flow field;
b) detecting discontinuities after the computation of the flow field;

c) simultaneous computation of the flow field and discontinuities.

In the first case, discontinuity detection is a task executed prior to normal flow being passed to a further stage for visual field computation. Global integration for computing visual field may be aided with motion boundaries obtained connecting motion discontinuities, thus optimizing the contours of the regions of integration. Discontinuity detection may only rely on normal flow information.

In the second case detection of motion discontinuities takes place after that computation of the visual flow field is done. This approach requires global integration in order to reconstruct the flow field from bare normal flow. Motion detection can work on full velocity field.

In the last case, the two processes work simultaneously and there could be some sort of co-operation between them.

2.6 Computing velocity field

Various methods have been developed for computing velocity field based on the normal flow information having in mind the biological feasibility of such algorithms.

Ulmann and Spoerri (1991) used the local histograms of the potential displacement to compute a dense image flow field using a well-posed method similar to the local voting scheme developed by Bülthoff, Little & Poggio (1989). In this method the discrete image flow field V(x,y) = (u(x,y),v(x,y)) (-/+μ,-/+μ) minimizes:

$$\int \Omega(E_t(x,y), E_{t+\Delta t}(x+u\Delta t, y+v\Delta t)) + \beta\left(\frac{d^2u}{dx^2}+\frac{d^2u}{dy^2}+\frac{d^2v}{dx^2}+\frac{d^2v}{dy^2}\right)dx\,dy$$

where $E_t(x,y)$ denotes the image brightness or intensity at (x,y), Ω is a comparison function which measures the pointwise match between subsequent frames and μ denotes the maximal expected displacement in the x and/or y dimension.

In Hildreth (1983) is described another method that takes into account a smoothness constraint, but that leads to an optimization problem as well.

The algorithm computes a velocity field solution that satisfies the constraints derived from the changing image and minimizes the measure of variation along contours given by

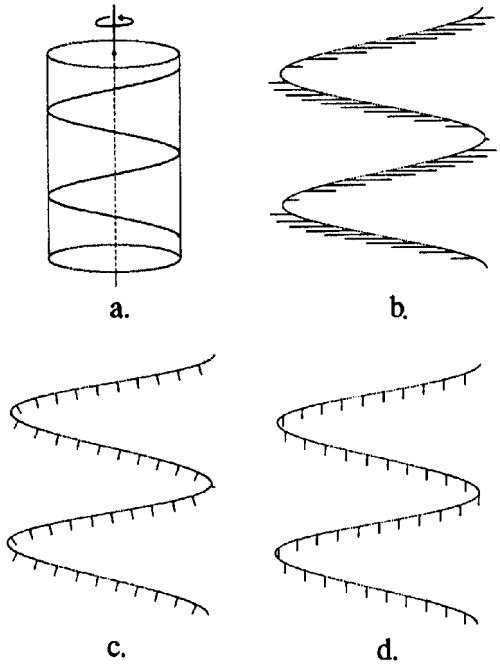
$\int \left|\frac{\partial V}{\partial s}\right|^2 ds$. Such approach leads to algorithms that involve simple, local and parallel operations that can be computed by a biological system.

The continuous functional, that leads to other discrete functions to be minimized in the complete algorithm, is the following:

$$\Theta = \int \left[\left(\frac{\partial V_x}{\partial s} \right)^2 + \left(\frac{\partial V_y}{\partial s} \right)^2 \right] ds + \beta \int \left[\boldsymbol{V} \cdot \boldsymbol{u}^{\perp} - v^{\perp} \right]^2 ds$$

where V_x and V_y are the x and y components of the computed velocity field, $\boldsymbol{V} \cdot \boldsymbol{u}^{\perp}$ is the normal component of the computed velocity field and $v^{\perp}$ is the measured perpendicular velocity component. β is a weighting factor that express our confidence in the measured velocity constraints. The optimization tends to select the computed velocity field that minimize the gap between the computed and the measured perpendicular components of velocity. The selected one is called **computed velocity field of least variation**.

The next illustration shows the computed velocity field of lest variation of the barberpole motion, that, as we know, is completely wrong and generates the famous illusion.



From Hildreth (1983): The Barberpole Illusion
a. the barberpole circular helix rotating about the vertical axis;
*b. the 2D projection of the helix and its **true** velocity field;*
c. the normal components of the velocity vectors in b.;
*d. the computed velocity field of least variation: in this case totally different from the true velocity field, hence the famous **illusion!***

Chapter 3 **Background**

In this section, relevant articles are discussed regard to the motion discontinuity issue.

3.1 Nakayama – Loomis model

Nakayama and Loomis (1974) proposed an hypothesis of how optical flow could be processed by relatively simple physiological mechanisms. They indicate the existence in the visual system of motion-sensitive cells that process the optical flow over the retina.

These cells should be organized in a center-surround receptive field structure where **C is the center region** and **S is the surrounding concentric region**. They should be directionally-selective, so that if the motion in a given direction *i* in the surrounding region is different from the motion in the central region, the cells fires up with the difference measured.

These cells could then be linked together in order to generate a higher-order cell sensible to a higher-order variable of the optical flow: the “convexity”. Convexity, under assumption of rigid movement, is related to relative depth. Such suggested cell is called convexity cell, has a center-surround structure, like the motion-sensitive cells, and is sensitive to the convexity function defined as follows:

$$C(\alpha,\beta)=\sum_i\left[\int_C V_i - k\int_S V_i\right]$$

where V_i refers to the component of optic flow in the direction determined by the value of *i*. The constant *k* takes into account the different areas of C and S so that the scalar function value (“convexity value”) is zero for uniform flow over C and S.

Since every motion cell is selective only for a given direction of the flow, the overall response of a given convexity cell at a specific location is the sum of all differences in each orientation. So, the convexity cell is sensitive to discontinuities of optical flow across the receptive field (independently of direction).

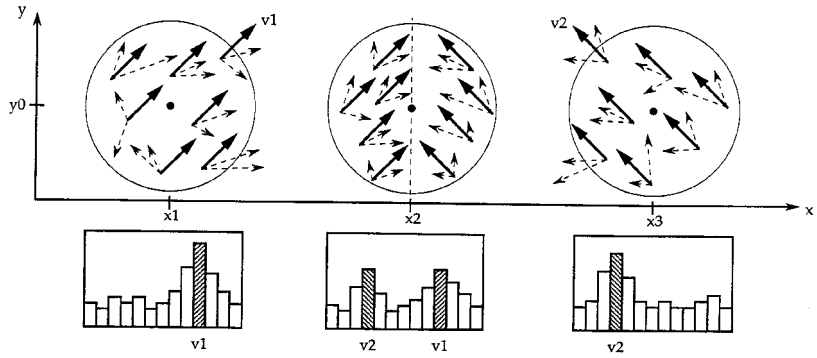
As evidence for the theory, they point out the existence of velocity-sensitive neurons that have inhibitory surrounds which are sensitive to movement stimulus and cells in the

monkey visual cortex which are preferentially activated from stimulus in their surround.

Personal considerations:
This model uses the center-surround schema and is the only one seen in this chapter. This key idea is biologically inspired and is the basis of this thesis implementation.
It's important to notice that the Nakayama-Loomis model uses optical flow as stimulus.

3.2 Spoerri thesis

Ullman (thesis supervisor) and Spoerri (1991) analyzed the early detection of motion boundaries proposing a 2-stage process based on motion information alone: (i) local estimation of motion discontinuities; (ii) extraction of complete boundaries of different moving objects.
For the first stage they developed 3 methods. Using potential displacements of an image point and the flow component normal to the intensity gradient, they developed a statistical model to analyze the local distribution of motion vector directions to look for bimodality present in the local histograms. A bimodal distribution indicated the presence of two local sub-regions with different motion directions. The Dynamic Occlusion Method, on the other side, computed locally the appearance and disappearance of thin bars, typically created or destroyed in in the vicinity of a motion boundary.



Spoerri (1991) - Bimodality in discontinuity detection
Examples of 3 situations where the histograms collect the potential displacements of the points that lie in the circle. Two peaks may lead to a motion discontinuity detection in the circle.

For the second stage, they modified the Structural Saliency Method to extract complete and unique boundaries from the pointwise output of the first stage, thus assigning a defined contour to moving objects in the scene.
In conclusion they argued that (i) useful segmentation of the scene can be performed on

the basis of motion information alone, (ii) estimation of motion boundaries can be decoupled from the computation of a fully image flow field and can be performed in parallel and (iii) proposed a method for extraction of salient, complete and unique contours of differently moving objects.

3.3 Koch's primate visual system motion model

Koch et al. (1989) adapts a gradient based computer algorithm for the estimation of visual motion to be computed by neurons in the primate visual system.

Given the time varying image intensity I(x,y,t) falling onto retina, the basic conservation law posits that dI/dt=0. Adding a smoothness constraint, the flow field is determined by minimizing a cost functional L:

$$L(\dot{x},\dot{y})=\iint\left\{\left[I_x\dot{x}+I_y\dot{y}+I_t\right]^2+\lambda\left[\left(\frac{\partial\dot{x}}{\partial x}\right)^2+\left(\frac{\partial\dot{x}}{\partial y}\right)^2+\left(\frac{\partial\dot{y}}{\partial x}\right)^2+\left(\frac{\partial\dot{y}}{\partial y}\right)^2\right]\right\}dxdy$$

It can be shown that the solution found for real images is qualitatively correct.

This is an area-based optical flow method, in contrast to the edge-based method proposed by Hildreth (see par. 2.6). The key idea is that the functional L to be minimized corresponds to power dissipation of a simple electrical network, thus the steady state voltage distribution corresponds to the minimum of L. Such network could be implemented using motion sensitive neurons of the mammals visual cortex, each one sensitive to a specific direction and orientation.

In the first processing stage local motion information is measured on a on-off direction selective basis, so that a direction of movement is computed for every location.

In the second stage the final global optical flow field is computed.

Though the algorithm respond well to a series of visual tests (perceptual phenomena and illusions), the major defiance of the methods using smoothness constraints is cutting out any discontinuities in the flow field. To counter this, is proposed (but not implemented) to use a Bayesian estimation and Markov random fields so that if the spatial gradient of the optical flow between two neighboring points is greater than some threshold, then a discontinuity is detected at that location and no smoothing is processed. Such an approach would segment different parts of the scene using motion.

Personal considerations

This is a first trial of implementation of a computer algorithm into a biological structure. Of course at that time a lot of knowledge was missing and today other ways are followed.

3.4 Grzywacz-Yuille model for local velocity estimation

Grzywacz-Yuille (1990) proposed model for local velocity computation using populations of motion sensitive cells. The reason behind this model is that motion-sensitive cells in the primary visual cortex are directionally-selective and tuned to spatio-temporal frequencies: these cells do not detect velocities, though humans can estimate velocities with high precision. They introduced a method to estimate local velocity from output of motion-energy filters that work correctly for pure translations and is consistent with cortical physiology. Other classes of motion, such as rotation or expansion, can be locally approximated to translation.

In order to compute velocity uniquely, they proved that it can be obtained by the largest responses of the motion-energy filters as a function of their optimal spatial frequency, optimal temporal frequency and optimal direction of motion. Grzywacz-Yuille presented 3 strategies for possible implementation with neuronal elements: ridge strategy (excitatory connections from each motion-energy cell to the velocity selective cell most consistent with it); estimation strategy (minimize a goodness-of-fit criterion to estimate the image's spatial characteristics and compute the velocity) and extra information strategy (uses the output of purely spatial frequency tuned cells to calculate the spatial characteristics of the image).

The model proposed is divided in two stages: the first stage measures motion energies from the moving stimuli and might take place in the primary visual cortex; the second stage estimates velocities locally from motion energies and might take place in the MT. The local computation could explain the phenomena of motion discontinuities and motion transparency (two superimposed planes moving at different velocities, so that for a specific location two different velocity vectors are defined).

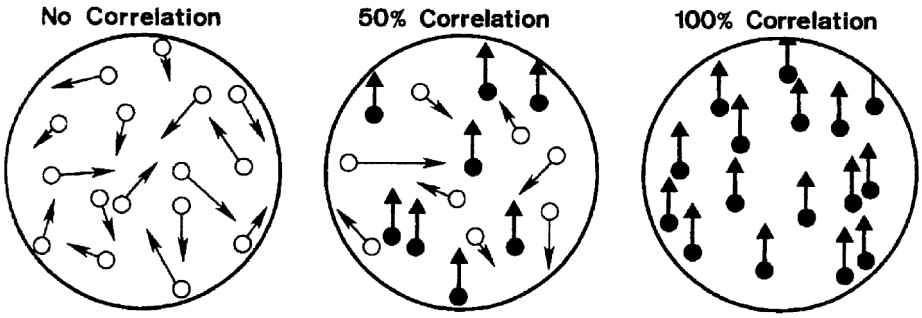
They also suggest that a third stage, that computes motion coherence through integration of motion over a global scale, could take place in a later cortical stage, not in the MT.

Finally, they argue that since the receptive field size of primary visual cortex cells is typically larger than regions of the visual world where texture exists, these cells have to deal with 2D patterns, not with the only gradient of luminance.

Personal considerations:
In this paper there's no aperture problem since they argue that MT is not concerned with it, but no evidence or argument is reported for this hypothesis. This is anyway one of the first biologically based models.

3.5 Newsome-Paré – Selective impairment induced by lesions in MT

Newsome and Paré (1988) induced lesions in the Middle Temporal Visual Area (MT) of two rhesus monkeys previously trained on psychophysical tests. MT area in primates computes many visual task, the experiments were conducted to analyze the effects of MT lesions in motion-sensitivity (with motion direction discrimination) and contrast-sensitivity (with orientation discrimination).
The visual motion-stimuli was a dynamic random dot display for motion correlation analysis (see following figure). Monkeys had to perceive coherent motion in the noise motion, setting the threshold at the minimum percentage of signal dots necessary to perceive coherent movement. The contrast-sensitivity stimuli was a stationary sine wave grating where monkeys had to discriminate the orientation.



Newsome-Paré (1988): visual stimuli
A sequence of dots is plotted by a computer, each dot survives 20-30 ms and then is replaced by another one at random location, so that it's impossible to track any dot movement. Net motion signal may vary between 0% and 100%

Lesions were made by injecting ibotenic acid into MT, a neurotoxin able to kill selectively cell bodies while leaving fibers of passage in the underlying white matter unharmed. Only one hemisphere was involved in the lesion, while the other one was left intact for control.
Results indicated a very high motion threshold rise after 24h from the injection (400-800%) in the hemisphere with lesion, while the contrast threshold had very little or none elevation. In the control hemisphere (with no lesion) both thresholds appeared completely normal and unchanged.

After 3 weeks from injection, the monkey's performance in the motion-sensitivity task improved considerably, but even at 5 months postlesion the motion threshold remained higher than prelesion, signifying a permanent deficit. These results indicate that MT lesions can produce permanent perceptual deficits, in part recoverable.
Histological studies where then performed to analyze the lesions induced.
The study conclusion was that MT plays a role in the selective perception of the motion, not only in its analysis.

3.6 Vaina et al. - Higher order motion tasks in patient with impaired motion mechanisms

Lucia M. Vaina et al. (1990) studied the motion perception of a 60 year old patient who had a stroke and reported a lesion in the extrastriate visual areas bilaterally, extending into the posterior parietal and temporal lobes (documented by MRI studies).
Background examination showed difficulty in touching objects in his reach, reading, writing and written calculations, but not oral calculations, meaning his difficulty was writing numbers and letters on the page. He was unable to judge lengths, bisect a line in the middle and copy a simple drawing. Ocular motility was normal. Contrast sensitivity, shape discrimination were normal. Binocular stereopsis and depth perception were impaired. Spontaneous speech, repetition, auditory comprehensions were intact; his verbal IQ was 104 but his performance IQ was 68.
Visual motion perception experiments showed that the patient successfully recognized a moving figure over a static random-dot pattern background, a notch in a vertical boundary generated by the movement of two random-dot pattern regions (except when direction difference was small), but was severely impaired in tests involving velocity magnitude differences (local speed discrimination). Motion coherence test, performed with a similar algorithm to the one used by Newsome-Paré (1988) described earlier, showed a significant impairment, indicating a difficulty in global motion integration.
Higher-order motion tests surprisingly showed and excellent capability of the patient to reconstruct a 3D rotating cylinder from dots movement and recognizing a human movement from the simple movement of lights attached to the joints (structure from motion reconstruction). Other patients with lesions to the right occipital-parietal area previously studied by prof. Vaina were completely unable to do this.
Results indicate that precise early motion measures are not necessary for higher-order

structure-from-motion tasks, thus invalidating a number of computational models proposed for human vision.

3.7 Vaina et al. - Deficits in local motion mechanisms

Lucia M. Vaina et al. (2003) studied the motion perception of AMG, a 53 year old patient who had a lesion in the left occipital lobe centered on visual areas V3 and V3A with underlying white matter involvement. The patient was tested for several motion task and results were contradicting several previous models that coupled the processes of motion integration and discontinuity detection.

The patient D-Max (i.e. the maximum displacement of the dots that can sustain a perception of a coherently moving array of dots) in the right visual field was smaller than the left one; the speed discrimination was impaired and a possible high (10-13 Hz) temporal frequency deficit was suggested by data. Thus, local motion mechanisms were found impaired.

Global motion tests surprisingly indicated that the patient was not impaired in motion coherence (a test where is determined the minimum percentage motion coherence at which a subject reliably discriminates the direction of coherent motion in a random-dot display) and was identical to normal control subjects.

In both the transparency and discontinuity detection tasks the patient was impaired in the right visual field but not in the left. In the discontinuity task the patient needed a percentage of signal dots four times higher than normal subjects, suggesting a deficit of integration across spatial scales.

The results indicated a specific impairment in the computation of local but not global motion and an inability to integrate motion information across different spatial scales. An impairment never reported before.

3.8 Rust-Mante-Simoncelli-Movshon MT direction selectivity model

Rust, Mante, Simoncelli and Movshon (2006) proposed a linear-nonlinear model to analyze the component direction selectivity and pattern direction selectivity properties of MT cells.

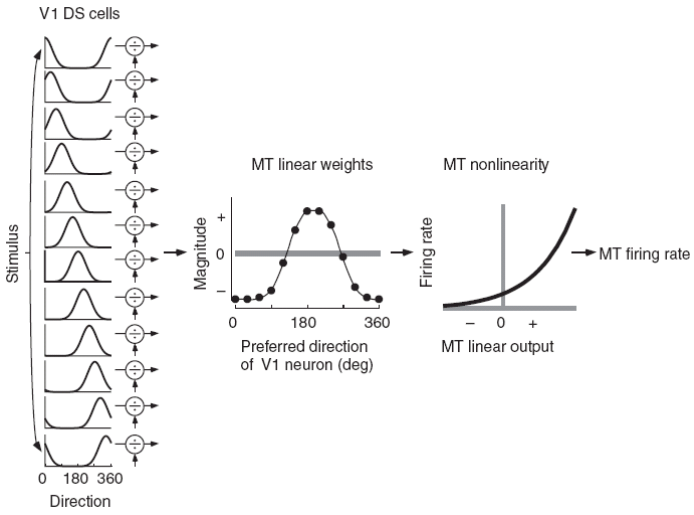
Another linear model for the local image representation using the properties of neurons in cortical areas V1 and MT was previously presented by Simoncelli-Heeger (1998), where the computation was performed in two stages (corresponding to V1 and MT) linearly

weighted and summed. Such model was both direction and speed selective, but failed to account for pattern direction selectivity.

It is known that MT neurons have non linear responses for single oriented gratings stimuli and for plaid stimuli obtained by superimposition of pair of gratings. There seems to be two MT cell types, even though evidence is equivocal: pattern direction and component direction selective cells.

The visual scene analysis requires elaboration of information represented by neurons in V1 (primary visual cortex) and direction selectivity in computed in V5 (extrastriate are MT).

They propose a cascade model where the stimulus is first passed through a population of 12 V1 direction selective model neurons with equally spaced preferred directions, then the MT model cell computes a linear weighted sum of such V1 responses (both positive-excitatory on negative-inhibitory). Finally the result of MT computation is non-linearly transformed into a firing sequence that simulates the actual cell response.



Cascade model: stimulus is processed by 12 V1 model neurons (Direction Selective cells) with direction preferences spaced by 30°, then the outputs are linearly combined with the linear weights of the MT cell. Finally the signal is transformed non-linearly into a firing rate of the MT cell.

Such model is then fitted with responses of individual MT neurons and it is shown that it reliably predicts responses to grating and plaids visual stimuli, capturing the full range of pattern motion selectivity found in MT.

They measured direction tuning curves for the responses of cells to gratings stimuli and to plaid stimuli, using adult anesthetized macaque monkeys. They recorded responses of 50

isolated direction-selective neurons in MT to visual stimuli presented at optimal spatial and temporal frequency within a circular window confined in the receptive field.

MT receives inputs from other areas besides V1 (like V2 and V3), but it seems that most or all nonstriate inputs to depend only on V1. Thus, only V1 is modeled. V1 stage is simulated using 12 model neurons with the well-known characteristics, but not including directional inhibition, adaptation, spatial integration and dynamical modulation (only steady-state response mechanisms were implemented).

The direction tuning curve of the model V1 neuron is described by a von Mises function:

$d_n(\theta_m) = e^{b*\cos(\theta_m - p_n)}$ where p_n is the direction preference (= 30*n deg)

Direction tuning curves are normalized to unit area and then the linear response of each V1 neuron to stimulus S is computed as:

$$L_n(t_i) = \sum_m d_n'(\theta_m) S(\theta_m, t_i)$$

V1 responses are the normalized in two different modalities: “tuned” and “untuned” which are combined by the following:

$$V_n(t_i) = \frac{L_n(t_i)^2}{\alpha_1 L_n(t_i)^2 + \frac{\alpha_2}{12}\sum_k L_k(t_i)^2 + \alpha_3 \overline{L}'}$$ where $\overline{L}'$ is the mean squared contrast of the hyperplaid stimuli.

Individual V1 responses are finally combined linearly into the response of the MT model cell $Q(t_i) = \sum_k w_k V_k(t_i)$ which is then converted non-linearly into a **firing rate** via a static non linear function $M(t_i) = f(Q(t_i))$

The authors suggest that such cascade models are accurate, computing parsimonious and may be useful in describing properties of sensory neurons far from the input stimulus.

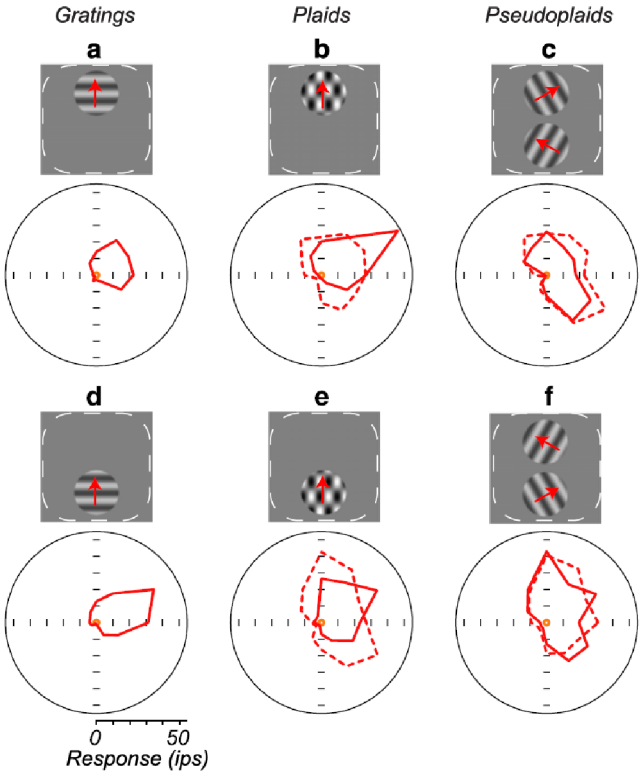
3.9 Majaj-Carandini-Movshon MT motion integration is local, not global

In their study, Majaj-Carandini-Movshon (2007) get to a rather surprising result concerning the scale of motion integration in the visual cortical area MT (V5), invalidating a number of previous visual models (like Simoncelli-Heeger (1998) model). Given that neurons in the

primary visual cortex (V1) have small receptive fields, that they are not able to perform any integration across space to obtain an estimate of the global object motion, and that they are component-direction selective cells, many studies identify MT as a candidate site for spatial motion integration. V1 cells only encode the motion of small local features, while MT cells receptive field is about 10 times larger than V1 cells and have two kind of direction selective cells:

1) component-direction selective cells – respond to component gratings matching their preferred direction of motion;
2) pattern-direction selective cells – respond to complex pattern and use information from overlapping components to compute the direction of movement.

In this study they tested if the computation of components/pattern movement involves the whole receptive field of a MT cell or just part of it. Using 12 anesthetized macaque monkeys, they recorded the responses of 54 MT cells to gratings, plaids and pseudoplaids stimuli as presented in the following picture.



Majaj-Carandini-Movshon (2007): polar plots of responses in spikes per second of macaque MT cells to gratings, plaids and pseudoplaids stimul. In this example,a plot of a pattern-direction selective neuron.

Stimuli were presented in two “patches” of 25-50% diameter of the MT cell receptive field, approximately equally responsive and placed along the axis of the cell's preferred direction.
The pseudoplaids in **c** and **f** are the gratings that, when overlapped with 120° direction difference, generate the plaids in **b** and **e**. Majaj-Carandini-Movshon idea is that if the computation of pattern motion were local, once separated the components of the pattern the neuron could not integrate the same motion; while if it were global, such separation should not compromise the result.

The 54 MT cells studied were divided into 3 categories, depending on their plaid-pattern response compared to the predicted response computed using the response to individual gratings:

1) 10 component-direction neurons: with a bi-lobed tuning curve with peaks corresponding to the preferred component direction of the two gratings that implement the plaid;
2) 20 pattern-direction neurons: they would only respond when the pattern moves in their preferred direction;
3) 24 unclassed: cells with a not well defined response.

This distribution changed using pseudoplaids: all of the cells that were component-selective in response to plaids, remained as such even in response to pseudoplaids, while all of the cells that were pattern-selective changed their behavior becoming unclassed or component-selective.
Majaj-Carandini-Movshon conclusion is that separating the components of a plaid into separate regions of the receptive field abolishes pattern motion selectivity in MT cells.
Among the hypothesis, pattern direction selectivity could begin earlier in visual areas such as V2 and V3, or perhaps MT can compute locally inside its receptive field, or again the computation is begun in V1 and then completed in MT. Segmentation of computation could be done by higher cortical areas with a feedback mechanism in MT under certain conditions. They conjecture that signals from MT only provide local motion measurements, which are integrated elsewhere with scene information to determine the final perception of coherent or incoherent motion.

3.10 McCool-Britten's review of Cortical Processing of Visual Motion

In this paper, McCool and Britten (2008) report the state-of-art knowledge about the cortical processing of visual motion in every step, from local to global. Having it summarized here provides a basic guide through the contents of this thesis with no intent of completeness.

Unlike color perception, motion perception requires computation, since all retina receives is a complex time/space-varying luminance pattern projection of a 3D world onto a 2D space, with occlusions, transparencies and other non-real situations. The question is how and where these computations take places, given the cortical hierarchy from local to global and the fact that cells only signal through firing rates.

This review is divided in three parts as follows.

1. Local Motion Mechanisms: V1

A primary distinction is set between Simple Cells and Linear Motion Mechanisms and a second type of Complex Cells and Non-Linear Motion Mechanisms. The first steps of motion processing is to compute local operations that detect the image contrast movement across space and time. In the first class of simple cells there are neurons with Receptive Fields (RFs) oriented in space and time, like direction-selective retinal ganglion cells and direction selective (DS) V1 cells in the geniculocortical pathway of the monkey. Such cells would respond preferentially to a particular orientation or direction of movement depending on the spatiotemporal profile of their RF. But there are also cells like the lateral geniculate nucleus (LGN) neuron that are nondirectional and nonorientation tuned.

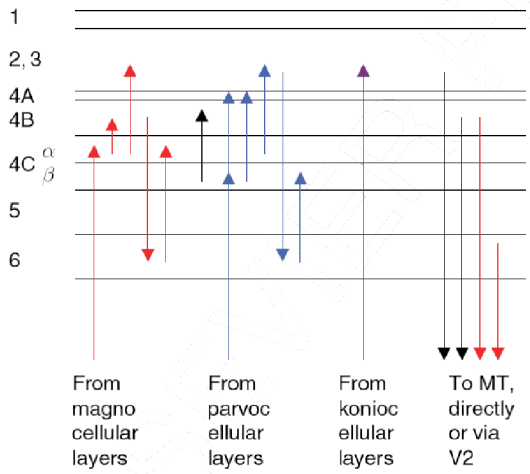
One result achieved is that there's a hierarchy through the primary visual cortex and that linear mechanisms can reproduce the directional responses for simple cells using superimposition, even with some underestimation in the magnitude of the direction.

Complex cells, on the other hand, produce nonlinearities that are not well described by simple models, like inhibition of responses in nonpreferred direction, multiplicative operations in preferred direction responses and the squaring operation, widely used in motion direction energy models.

These cells don't have distinct RF subregions and so are insensitive to the phase or the location of the stimulus. Local responses are combined nonlinearly and RFs have a second-order non linear profile oriented spatiotemporaly matching the cell's preferred direction and speed. For example, direction selectivity can be computed using a motion

energy model that squares and then sums the output of two RF filters.

The neuronal circuitry is divided per areas and these areas are interconnected in many ways. It is possible to set a structure like the one in the picture:



McCool-Britten (2008): segregated input in V1 from magnocellular, parvocellular and koniocellular layers

Input in the V1 is segregated and comes from the magnocellular layers of the LGN (like parasol ganglion cells that are temporally acute and highly sensitive to contrast), the parvocellular layers and the koniocellular layers. Magnocellular information can enter via a monosynaptic connection between layer 4Cα and 4B and, together with color and form sensitive parvocellular cells, in layer 2 and 3. Large pyramidal cells in layer 4B receive input from magnocellular and parvocellular and project to MT area. Magnocellular cells project also to layer 6 that is one of the more directionally selective layes, having large Meynert cells that can sum directional information across space.

Complex cells are found primarily in layers 4B, 2/3, 5 and 6 and direction selectivity is computed in upper layer 4 and 6. Considerations from tests are that V1 input in MT is still largely unmixed, though object contours and directions are primarily calculated in V1 directionally selective neurons. Another 15% of V2 cells and 40% of V3 cells are direction selective, where V2 receives 67% of its input from V1 and V3 receives both magnocellular and parvocellular inputs. V3 cells also respond to plaid stimulus, like MT. Finally, both V2 and V3 receive strong feedback from MT probably to modify their analysis of moving stimuli (luminance or contrast).

The spatial scale of the RFs approximately doubles with the level over V1, starting from about 15 arcmin for V1 RF near the fovea, twice in V2, four times in V3 and 8-10 times in MT. The frequency domain properties of the detectors are clear, like the behavior of reducing nonpreferred temporal frequencies.

2. Medium Scale Motion Processing: Area MT

Middle Temporal (MT) Area – or V5 – is located in the middle of the dorsal stream and is characterized from having 80-90% of its neurons strongly directionally selective, firing when motion matches their preferred direction.

Its input comes mostly from the layer 4B of V1, but also from V2, V3, V3A, VP. V3d and PIP and is highly specialized with large diameter axons forming multiple synapses onto MT neurons. Some neurons show inhibition for motion in the direction opposite to the preferred and some others are selective for binocular disparity. Since directional computation in the MT neuron occurs on as scale much smaller than its receptive field, it has been suggested that MT inherits its directionality from earlier computations such as V1; other studies reported MT firing rates correlated with speed perception and acceleration.

Integration in MT has many problems, beginning with the aperture problem: since the component of motion parallel to the edge is invisible, it is possible to estimate a single perpendicular direction. Thus, in order to obtain an estimation of the actual velocity of a moving object it is necessary to integrate all the single vectors of the moving contours.

20% of the MT neurons are selective to the plaid stimuli direction, even though the components of the plaid stimulus are not moving into that direction, meaning that these neurons are integrating the motion vectors.

It has been proposed that, alternatively, a subset of V1 cells, “end-stopped cells” that respond only to the endpoints of the contour, could transmit the true motion vector of the plaid stimulus to the MT (since that point is the only one moving in the plaid direction).

Another function is the speed perception and estimation. This is done through correlation of spatial and temporal frequencies of the moving object perceived by V1 populations. It is not clear how much MT integrates and what is done in V1, but frequencies bandwidths are broader in MT and some MT cells show an activation for a single preferred velocity.

A notable problem for integration is the segmentation of the different objects moving in the

scene, since signals coming from the same object have to be grouped before being integrated. To solve this, about half of the MT cells have an antagonistic surround mechanism, where a surround area -external to the receptive field- if activated suppresses the neuron's response for stimuli in the cell's preferred direction. Similarly, reinforcing surrounds enhance the cell's response. Color is another way to segment objects.

Contrast is another problem for integration, since it is known that a low contrast of the visual stimulus decreases the perception of speed and the firing rates responses of neurons. In V1 reducing contrast lowers preferred spatial and temporal frequency.

A lot of studies were conducted to test whether MT is the only cortical area responsible for motion perception, but while a lot of evidence strongly correlates MT neuronal spiking activity to performances in perception tests, large MT lesions do not completely cancel these abilities. Besides that, it is shown that MT is not completely specialized in motion perception.

Another studied property is the neuronal adaptation to preceding stimuli. Motion adaptation is well seen in the motion aftereffect (MAE), where after adaptation to a moving stimulus, motion is perceived even if nothing is on the retina. Motion adaptation can be local (evidence of directionally selective adaptation in V1) or global (adaptation to complex patterns), but there's no proof that MT is critical for adaptation, although is an important locus.

Many experiments showed that cognitive contributions, such as attention and memory, may modify the MT neuronal properties. Attention can select important areas in the scene and can be independent of the eye fixation location. It can also modestly shift the MT cell receptive field and modulate its firing rate. Memory of the motion direction can also modulate the MT firing rate, but there's no knowledge of how these cognitive controls operate or where they come from.

MT neurons can be studied in detail only in animals, while in humans can be used only the low spatial and temporal resolution functional magnetic resonance (fMRI) to measure the neuronal activity. Despite this, it has been found that human and macaque MT are functionally similar and respond depending on the coherence and contrast of the motion signal. Human MT is activated by first and second order motion and is more sensitive than the macaque; it can be activated even by mental imagery of motion (specifically rotation), without any visual motion.

3. Global Motion

Global motion is the whole motion pattern result of the world around us. Its projection onto retina is called **Optic Flow** or retinal flow. It incorporates the movements of the head and the eyes, so that the vector pattern produced by these self-movements is superimposed to the vector pattern produced by the global motion.

Forward motion causes an expansion movement generated from a point called focus of expansion (heading). If objects overlap, there are depth discontinuities that generate sharp changes in the velocity pattern on either side of the boundary.

Medial Superior Temporal (MST) area is the most studied for global motion. It is connected to MT and receives strong feed-forward input. MST is conventionally divided in distal subdivision (MSTd), responsive for very large stimuli, and lateral subdivision (MSTl), sensitive to small stimuli.

MST receptive fields are very large and could cover the entire visual field, while MST responses are mostly selective for direction and complex motion patterns like expansions, contractions and rotations (or combinations, like spirals). For example, a neuron could be sensitive to clockwise contracting spirals or changes in heading direction.

MST also receives extraretinal inputs like position and velocity of the eyes and vestibular signals. It is known that MST plays main a role in generating pursuit eyes movements and thus visual information could be a feedback for correction of pursuit. In addition to that, extraretinal information could help to stabilize the heading representation distorted by eye and head movements, differentiating real world motion and self-motion. Vestibular system signals to MST linear and angular accelerations, thus helping to compensate head rotations and amplify heading signals.

MST is involved in the discrimination of pattern motion of medium scale (10-40°), thus contributing to motion perception too.

Other areas responsive to optic flow patterns include Area 7A, selectively responsive to radial motions. Area ventral intraparietal VIP, that projects onto area 7A and is tuned for expansions, is capable of heading encoding. VIP responds to visual, vestibular and tactile stimuli and is believed to have a role in encoding object motion in near-extrapersonal space. Motor cortex , area 5, area PEc (dorsal pathway), area STPa (ventral pathway) are also responsive to optic flow stimuli. Area STPa is selective for object motion.

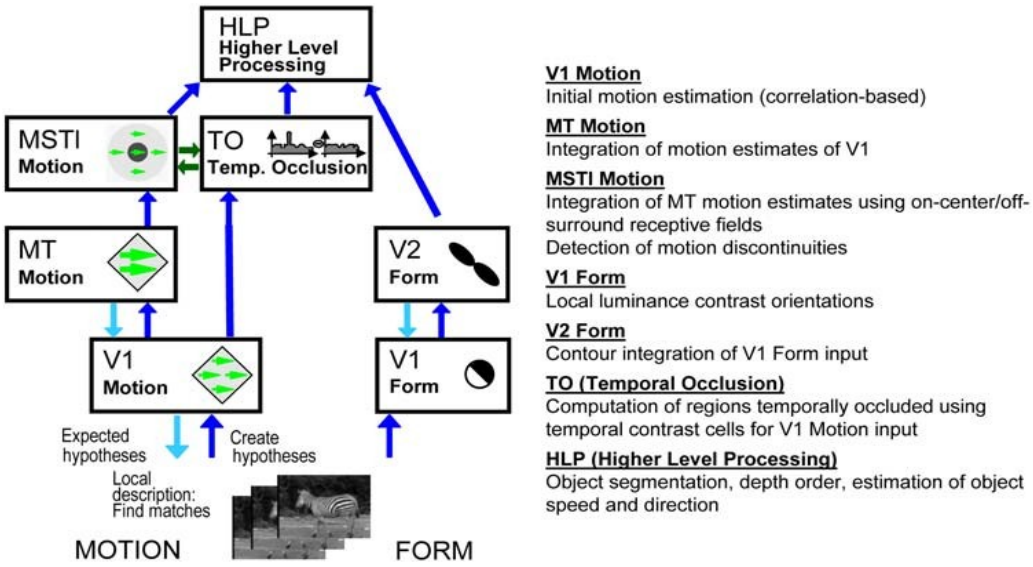
McCool and Britten conclusion is that all the synaptics of direction selectivity is a mystery, like all circuitry above V1 and most of the structure-function relations of the motion system. It is well known what happens, but with no certainties on how this happens. They blame in part for this situation the lack of data to support biologically realistic model.

Personal considerations:
Quite surprisingly, motion discontinuity or boundary detection problem is never cited nor reported in this paper. This could be symptomatic about the lack of responses in this field and thus an hint to go on with more studies in this crucial area of motion perception.
The use of V1 end-stopped cells instead of integration in MT, suggested in the aperture problem discussion, could be compatible with the findings of Majaj-Carandini-Movshon (2007) reported earlier in this section of the thesis.

3.11 Beck-Ognibeni-Neumann biologically inspired model

Beck, Ognibeni and Neumann (2008) developed a biologically inspired architecture that integrates information of different model components of the visual processing using optic flow. The purpose of their work is to obtain a model that implements object segmentation (i.e. detection all boundaries of a moving object) using only kinetic boundaries.
Object segmentation is obtained using both motion discontinuity and occlusion detection with temporal integration. An “occlusion” happens when an object covers another object or the background. As the object moves, parts of the background texure is covered, while other parts are disoccluded. This kind of information can be used to detect the movement and the boundaries of the object.
The computation of motion discontinuities is based on spatial contrast detection, while the computation of occlusion regions is based on temporal detection.

As the picture shows, there are a lot of feedforward and feedback connections between the various modules. The left chain computes motion, while the right computes form.



Beck-Ognibeni-Neumann (2008): A biologically inspired model for object segmentation from motion discontinuities and temporal occlusions

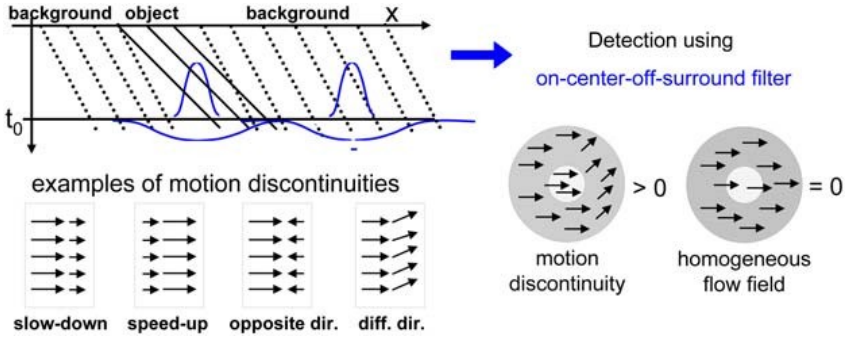
For the motion computation, in the primary visual area V1 stimuli are analyzed in parallel for motion direction and then projected to MT. In MT, where specific neurons exist, are computed direction and speed of the 2D image velocity through integration. The computed optic flow is the passed to MSTI that detects object motion using special units with center-surround motion fields. MSTI interact with the component for the detection of temporal occlusion, TO (not linked to any specific cortical area). TO is fed with the V1 initial motion detection.

The form is computed using a feedback/feedforward combination of V1 directed-contrast sensitive neurons and V2 long-range filter neurons that group contours.

All MSTI, TO and V2 signals converge to HLP (Higher Level Processing) component, not linked to any specific cortical area, that integrates them into an interpretation of the scene with segmentation of the image and ordinal depth order of the objects.

Optic flow detection is computed integrating in MT the raw and noisy estimates from V1 cells, though reducing spatial accuracy. Receptive field of V1 and MT are with ratio 1:5.

Particularly relevant for our matter is how motion discontinuities are detected. The authors model a motion discontinuity detector with an on-center-off-surround receptive field that respond very strongly if center motion and surround motion differ. This detector neurons would be located in MSTI and receive input from MT neurons.



Beck-Ognibeni-Neumann (2008): Motion discontinuity detection with some examples and a typical on-center-off-surround filter response

Spatial integration: for each position is calculated the mean velocity. The flow vector at position **x** is:

$$\overline{v_x} = \left(\sum_{all\ neurons\ at\ \mathbf{x}} u_{\mathbf{x}}^{MT} \cdot v_x \,,\ \sum_{all\ neurons\ at\ \mathbf{x}} u_{\mathbf{x}}^{MT} \cdot v_y \right)$$

where $u_{\mathbf{x}}^{MT}$ is a weight of the MT activity in **x** and (v_x , v_y) are estimated velocities.

If the mean velocity at a surround position is similar to the mean velocity in the center, the neuron activity is inhibited, while the positive activity is integrated in time to stabilize the MSTl model and added to the current motion discontinuity value with decreasing weight in time.

Spatial contrast responses $w_{\mathbf{x}\Delta v}^{MSTl}$ are computed as follows:

$$\delta_t w_{\mathbf{x}\bar{v}}^{MSTl} = -A\, w_{\mathbf{x}\bar{v}}^{MSTl} + B \cdot \left(u_{\mathbf{x}\bar{v}}^{MT} - \sum_{\mathbf{x}'\bar{v}'} u_{\mathbf{x}'\bar{v}'}^{MT} \cdot \Lambda_{\mathbf{x}\mathbf{x}'}^{S} \right)$$

Results are then grouped to derive a segmentation of the scene based on the motion discontinuities.

Personal considerations:

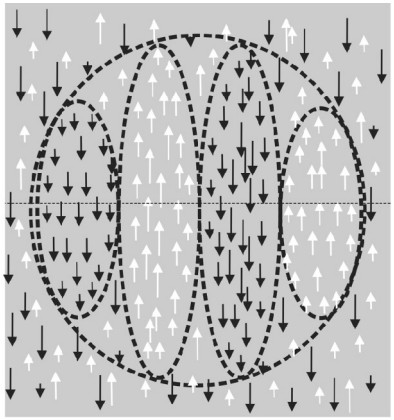
As a final comment, this work uses optic flow for its computations, that we know is not available at early stage from V1 processing due to the aperture problem. It seems that this model requires too much information to properly compute the scene, while we know that motion discontinuity detection is performed at very first stages, prior to global integration. They also indicate MSTl as locus for a motion discontinuity detectors fed by MT, while it could well be before that.

The model proposed uses on-center-off-surround receptive field detectors, like our model, and computes correctly a number of real and simulated situations, hence it could be modified in order to use normal flow instead of optic flow.

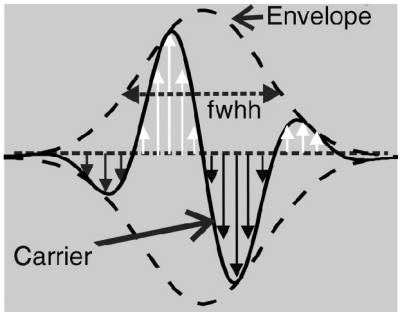
3.12 Durant – Zanker motion contour detection

Durant-Zanker (2009) in a recent study tested the motion contour detection in humans using a novel stimulus based on a 2D Gabor function. Two main kinds of experiments where performed:

- Motion-defined patterns: in the first series a motion-defined Gabor pattern of very small black moving dots was presented to volunteers and they had to detect the correct orientation of the Gabor stimulus (randomly chosen);
- Luminance-defined patterns: in the second series, luminance replaced motion in the Gabor stimulus, so that the Gabor pattern determined the luminance of the dots ranging from black to white (in a light gray background) instead of their motion;



Durant-Zanker (2009): dots velocities in the 2D Gabor stimulus (vertical in this case)



Durant-Zanker (2009): shape of velocity profile across space. Full width at half height (fwhh) is shown.

The stimulus was composed by 0,05° dots moving at a speed of maximum 3 pixel/frame (=10°/s) with a maximum lifetime of 50 ms, randomly located in a 12.5°x12.5° square area around a fixation target, presented on a 21" CRT monitor observed at 57 cm from the screen for usually 15 frames (250 ms) at a resolution of 656x493 pixels.

The stimulus duration was altered from a minimum of 1 to a maximum of 60 frames using a standard up-down staircase procedure, according to whether the response was correct.

The results of the experiments showed that the detection improved with the increasing envelope size and leveled off at around **4°-5°** full width at half height (= about 8.5° receptive field size) and decreases with higher spatial frequency of the Gabor pattern, with best performance at 0.1 cycles/degree.

This suggests that motion-defined contours are integrated on a relatively large scale (8.5°) and that detectors responding to changes in the motion field are specialized for detecting

motion edges rather than being frequency analyzers for reconstruction of texture surface motion.

Finally, motion-defined Gabor patterns and sparsely defined luminance Gabor patterns produced similar results at low sampling frequencies.

In the discussion the authors cite studies where fMRI on humans and macaques do not clarify the areas involved in kinetic contours extraction and thus they do not make any hypothesis about where the function is computed (MT, MST, IT, KO-V3B, V4, V3) but they report that V1 and V2 had some evidence, though controversial.

Particularly interesting is the size of the optimal stimulus they found (4°-5° full width at half height) that is quite large.

Chapter 4 **Methods: the algorithm**

Nakayama and Loomis (1974) proposed a center-surround model for identifying locations on the image plane where there was a considerable local variation of the motion vector directions and they argued that this idea is easily extended to include speed.
Our implementation is based on a modified version of the Nakayama-Loomis model. The modification was necessary to be able to use **normal flow** (impoverished version of the optic flow used by Nakayama-Loomis).
The goal of the local model is to detect motion discontinuity based on measurements within a small aperture, comparable in size to a typical convolution kernel.

4.1 Motivations for a local model in motion discontinuity detection

The search for a local model was motivated by double dissociation between motion coherence and motion discontinuity detection that Prof. Vaina found in patients.
The double dissociation suggested that **coherence** -which requires global integration- and **discontinuity** are perhaps not computed simultaneously (nor is discontinuity computed at a stage that follows coherence computation). A notable candidate to explain the double dissociation is a *local* model for discontinuity detection that operates independently of the global integration required for motion coherence.
The motivations to the use of normal flow in discontinuity measurement are the following:

1. normal flow is computed locally;
2. normal flow computation is not affected by the presence of discontinuities (in contrast to full optic flow computation);
3. normal flow can be computed non-iteratively.

The first motivation was discussed in § 2.3, the second one comes naturally because since computation is local, there are no possible discontinuities. The third motivation comes from the fact that each normal velocity vector is computed independently form the others, thus the whole normal flow computation can be parallelized.

4.2 Nakayama-Loomis model for detecting motion discontinuity

The Nakayama-Loomis model is a center-surround mechanism. They suggested a **"convexity" function** that assigns to each pixel location a scalar value which is determined by the optic flow over a center region (C) and a concentric surrounding region (S):

$$C(\alpha,\beta)=\sum_i\left[\int_C V_i - k\int_S V_i\right]$$

where V_i refers to the component of optic flow in the direction determined by the value of i. The constant k takes into account the different areas of C and S so that the scalar function value ("convexity value") is zero for uniform flow over C and S.

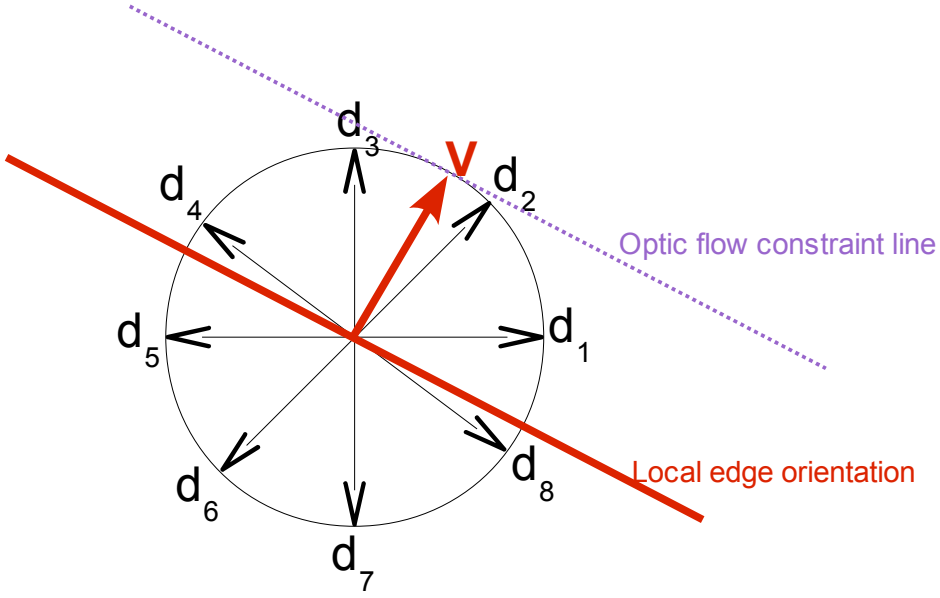
The scalar function will have a high value at discontinuities and low values elsewhere. By thresholding, the discontinuities can be isolated.

4.3 Extension of the Nakayama-Loomis model

We have extended the above model for situations where only a local projection of the optic flow is available due to the aperture problem (see specific paragraph 2.2). The local projection, termed normal flow, can be computed easily from two or more frames of an image sequence.

Since only the component of optic flow along the local intensity gradient direction is available, to address this information loss we extended the Nakayama-Loomis model in the following way.

In the Nakayama-Loomis model, for each direction considered (indexed by i in the above equation) the projection of optic flow vector in that direction is used in calculating the value of the convexity function. However, since optic flow is not computed in our model, we advised a **voting scheme** instead of the integral in the above equation. In this voting scheme, every normal flow vector votes to a set of directions as shown schematically in the figure below:



In the figure, the normal flow vector **V** is oriented orthogonal to the local edge orientation. The optic flow constraint line is the locus of the tip of all possible optic flow vectors corresponding to the normal flow vector **V**. Clearly, all of these vector will have directions within a 90 degree range on either side of **V**, but their magnitude will depend on the optic flow constraint line. In other word, any of the possible optic flow vectors will have components in the directions d_1 to d_4 for the example shown.

However, without knowing which is the correct optic flow vector, it is impossible to calculate the magnitude of these components. Given this limitation, we chose to treat all components equally. Thus, all components within a 90 degree range of the normal flow vector get a vote each. To compute the convexity function, for each direction d_i we add up the votes within the central region C (N_i^C) and subtract *k* times the total votes within the surrounding region S (N_i^S). After squaring this difference, we sum the result for all the directions considered:

$$\tilde{C}(\alpha,\beta)=\sum_i \left[N_i^C - k\cdot N_i^S\right]^2$$

The resulting approximation to the **convexity function** of Nakayama and Loomis is then thresholded to find location with significant motion discontinuity. A neuron computing this modified convexity function over its receptive field would be in fact a "**motion**

discontinuity detector" and can be called "**Convexity cell**" (term used in Nakayama and Loomis, 1974).
As in the original model, many direction-selective velocity cells with center-surround receptive field, centered in the same retinal locus, converge into a single higher order convexity cell that sums the votes received.

4.4 Biological and physiological motivations for the extended model

Direction-selective Retinal Ganglion Cells are selective for normal flow, since they suffer from the aperture problem, hence **they can only detect the normal component** of the movement that activates them. This is consistent with the original hypothesis of using bare normal flow for motion discontinuity computation.

The ideal location where this algorithm could be computed in primates is neurons in primary visual cortex. V1 neurons receive input directly from Retinal Ganglion Cells (RGC), through the lateral geniculate nuclei (LGN), thus they can receive normal flow detected by RGC directly. They have local reception field size of variable from 15 arcmin for V1 RF in the fovea (McCool-Britten 2007) to some degrees (Cavanaugh-Bair-Movshon 2002), are known to implement a set of selective spatiotemporal filters and in particular some of them are direction selective, i. e. respond strongly to a preferred direction of movement, or speed selective, i. e. respond strongly to a preferred speed of movement.
The local information coded in V1 neurons is based on local contrast, rather than brightness, and this is consistent with our hypothesis.

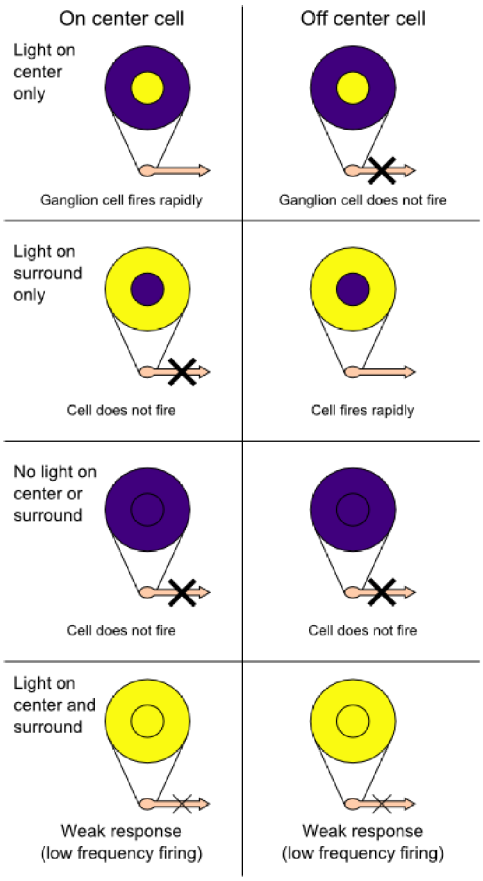
On the other side, visual area Middle Temporal MT (or V5) is known to receive local input from lower visual areas and integrate middle scale motion processes. MT neurons have higher receptive field sizes. **Rough motion boundaries computed by V1 neurons** with this algorithm could be a **first plot** where to integrate signals, to be later cleared and defined with precision into well defined motion discontinuities.
A support to this theory is found Majaj-Carandini-Movshon (2007) where it's proposed that direction selectivity could begin earlier than MT, like in V2 and V3, or begun in V1 and completed in MT. In our case the algorithm uses a voting scheme but it's based on direction selectivity.

MT cells respond with a non-linear firing rate (Rust-Mante-Simoncelli-Movshon 2006), a further step for this model would be using this property to filter the input from V1 cells into an enhanced on/off pattern of motion boundaries. On the other hand, V1 output is linear (McCool-Britten 2008) and thus our model is consistent.

Medial Superior Temporal area (MST) integrates MT signals into a global motion pattern, with receptive field sizes varying from 10° to the entire visual field (McCool-Britten 2008).

The proposed algorithm uses basically **On-center-Off-surround receptive fields**, that we know to be present in many neurons.

In Beck-Ognibeni-Neumann (2008) a very similar approach is used for computing motion discontinuities. They use in their model a MSTl neuron with a on-center off-surround receptive field that will respond very strongly if center and motion surround differ ("motion discontinuity detector"). But their input comes from the integration of MT, while our input is just normal flow (V1). Since MST has connections with V2 and thus with V1, it's plausible that Beck-Ognibeni-Neumann model could work with our algorithm using local information.



Representation of On/Off-center retinal ganglion response (from Wikipedia at http://en.wikipedia.org/wiki/Receptive_field)

Finally, Durant-Zanker (2009) experiments showed that motion contours are processed on a quite large scale that saturates at about 4-5° full width at half height of the Gabor stimuli used (8.5° total receptive area). This is a useful track about the size of the receptive field of our hypothetical Convexity cell.

4.5 Input of the model

4.5.1 BRAVI tests

The tests performed at BRAVI Lab. (Brain and Vision Research Laboratory, Department of Biomedical Engineering, Boston University) are based on the algorithms described in Vaina, Lamay, Bienfang, Choi and Nakayama (1990); Vaina, Gryzwacz, Saiviroonporn,

LeMay, Bienfang, Cowey A. (2003).

Motion coherence stimuli is a dynamic sparse random-dot kinematograms generated by a PC and displayed on a monitor in an circular **aperture of 10° diameter (79 dg^2)**. Each dot has a defined probability of being signal or noise, the strength of the signal is the percentage of correlated dots moving in the same direction (0%-100%).

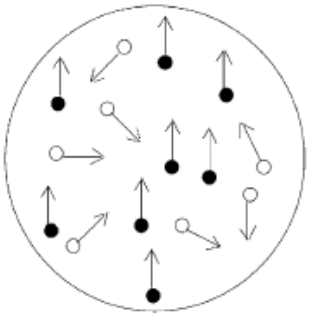
Random dots are used to minimize familiar position cues and to isolate motion mechanisms. Each dot in the aperture has an equal probability to be paired with a correlated dot in the subsequent frame and thus to contribute to the total motion signal. The correlated partners in turn have the same probability to be succeeded by another correlated partner. So, if the probability of being signal is 0.1 (10%), the probability for a dot to continue the same path in **3 consecutive frames is 0.001 (0.1%)**, making very unlikely that an observer tracks any dot's movement or any local cluster of dots. The size of the step of the dots is constant at 9 arcmin, the speed of motion is 3 deg/s. A conventional “wrap-around” scheme was used, in witch dots displaced beyond the aperture reappeared on the opposite side of the aperture.

Each test lasts 1 sec and are showed 22 frames, thus the frame's life is 45 ms (optimal for psychophysical experiments with humans and monkeys).

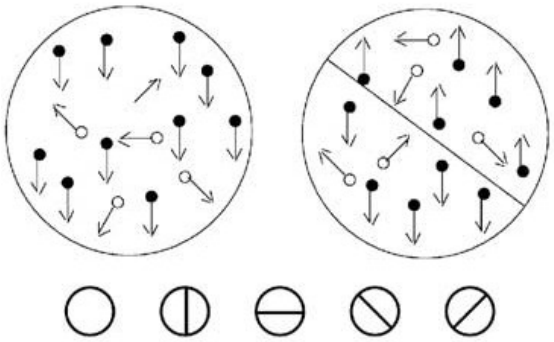
Dot density is 2 dot/dg^2, thus **in a 79 dg^2 aperture there are 160 dots**.

The patient is instructed to maintain fixation in the center of the circle where lies a mark. When viewing this stimuli at 50% signal, **the overall impression is of a twinkling visual noise with a motion signal embedded**.

In a **motion coherence test** the aim is to determine the threshold of motion signal necessary to correctly discriminate the global direction of the motion (upward, downward, etc...) in a series of tests where coherence percentage is progressively lowered by an adaptive staircase procedure.



In a **motion discontinuity detection test** the aim is to determine the threshold of motion signal necessary to reliably detect discontinuity from motion cues in a stimuli where, at random, an imaginary line divides the display in two parts, defined by the motion of the signal dots in opposite directions. Even in these tests coherence percentage (% of net motion signal) is progressively lowered by an adaptive staircase procedure.



The possibilities are shown in the bottom:

- no motion discontinuity – signal dots move all in the same direction;
- vertical discontinuity
- horizontal discontinuity
- -45° discontinuity
- +45° discontinuity

4.5.2 Model's coded stimuli

The key idea of this implementation is to simulate the stimuli described earlier coding an input directly into the model. Since this thesis purpose is to understand if the model is correct in it's answer to a well-known stimulus, we didn't implement the whole part regarding the scanning of the retina in order to acquire two frames of the moving stimuli and then calculate the instant normal velocity vectors.

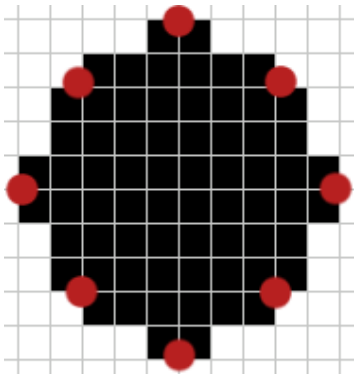
We coded directly the normal flow input into the model, thus knowing we have perfect conditions and not real ones.

Since the stimulus is made of equal dots, we modeled one dot as follows.

The **dot diameter** was assumed to be **10 minutes of arc** (10' or 10 arcmin), see Beardsley and Vaina (2001), and **visual acuity equal to 1 arcmin**. This assumption on visual acuity (for humans) is confirmed by Ganong (2006, p.162). Visual acuity is the

minimum grade of detail and object contour that can be perceived. Thus, the dot can be identified with a 10x10 matrix. That's the “resolution” of the retina, so that's what physiologically retina can perceive.

We considered 8 significant and conventional locations in the dot, identified in red in the following picture.



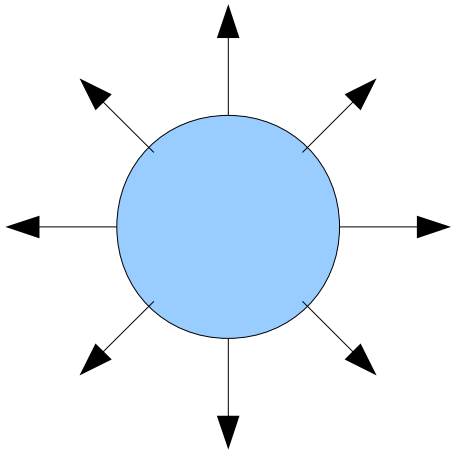
In these locations, named “**subdots**”, lies the local gradient of light intensity where normal velocity vector is computed. They are the “**edge**”, the border of the dot.

In our model, every moving dot generates 8 subdots movements too. Since for the aperture problem only normal component of the velocity vector can be computed at an edge, given the true velocity vector of the dot we calculated the **normal velocity vector** for the 8 edges of the dot. The normal velocity vector is obtained **projecting the true velocity vector onto the normal direction to the edge**.

With this simplification, only 8 directions are possible in our model, but we think they're enough to give sufficient data to have a good qualitative idea of the results. An improvement of the model could be upgrading it to 16 or 32 subdots/directions.

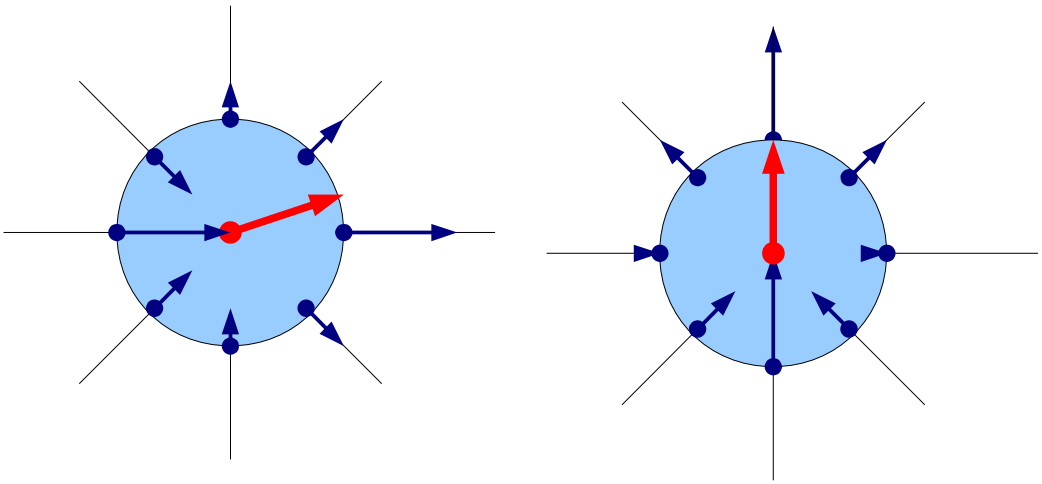
$$\vec{v}=\alpha_1\cdot\vec{d}_1+\alpha_2\cdot\vec{d}_2+\ldots+\alpha_n\cdot\vec{d}_n=\sum_{i=1}^{n}\alpha_i\cdot\vec{d}_i$$

Where v is the true velocity vector, α_i are scalars, d_i are the possible direction vectors and n is the number of locations considered for subdots, 8 in this case.



8 locations of detection ("subdots" or "edges") = 8 possible directions in our model

Here follow two examples of how normal velocity vectors of “edges” (blue vectors) are computed given a true velocity vector (red vector in the center of the light blue dot).



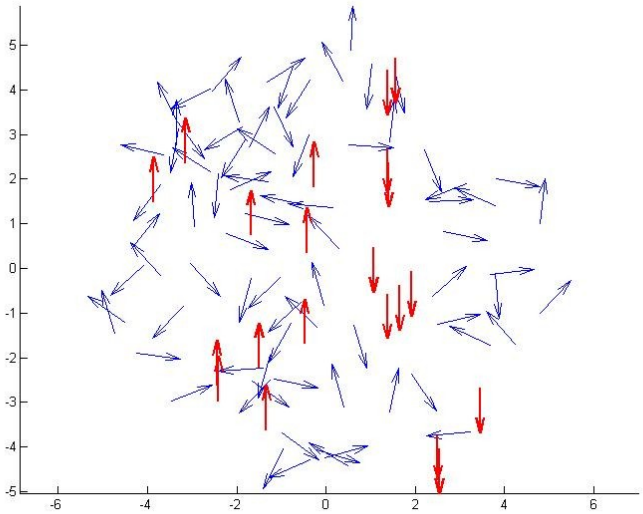
Another simplification is that two dots could superimpose during their short movement, since the initial location is chosen randomly, thus occluding part of the dots. In this case a dot wouldn't generate all 8 edge's normal flow vectors, but only a part. Anyway we considered this phenomena irrelevant respect to the total number of normal flow vectors, since it's a rare situation and didn't require a specific treatment. An evolution of the model could take account of it.

In the real tests, **dot's life is just few milliseconds**, to avoid eye tracking of the single dots, and when a dot dies another one is generated at a random location (see previous paragraph for more details). This permits only a global movement perception.
The whole test lasts 1 second and in each of the 22 frames there are 160 dots. That means that **more than 3.500 dots are showed** to the observer during the stimuli exposure.
We coded this situation into our model **compressing the total 22 frames test in one single velocity flow frame**, thus including a sort of temporal-integration during the experiment. In this frame are recorded all data about location and true velocity vector of the various dots printed on screen. To do this, it's important to remind that in order to calculate a velocity, 2 frames are needed: that means that in one session of the test can be calculated 160 x 11 = **1.760 dots velocities**. Since is very unlikely for a dot to last

more than two frames (to avoid eye tracking), we assume that all dots disappear after two frames.
The input is the true velocity flow, but then the algorithm of the model calculates the normal flow and uses it for its further computations. Complete MATLAB code of the algorithm is reported in Appendix.

The center-surround receptive fields are considered circular as assumption. Although real profiles do vary, this simplification is not unreasonable.
From Cavanaugh-Bair-Movshon (2002) studies on Macaque V1 neurons, we know that surround mechanism is 2.5 times the width of the center mechanism, thus the fields extents of Center and Surround are similar. From this study we found that center width average is **1.4°** and surround width is **2.7°**, for small eccentricities (less than 5°). We took this values as reference for our simulations.

In the next picture is reported a situation of a simulated motion discontinuity test with 100 moving dots, 20% of signal (red) – 80% of noise (blue), and a vertical motion discontinuity. Signal dots are in red for identification, but on real test there's no difference at all.

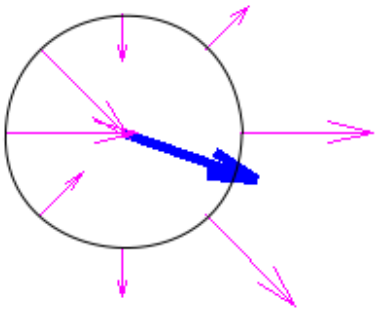


A vertical motion discontinuity: 20% motion signal (bold arrows in red) describes a vertical discontinuity with another 80% of motion noise.

4.6 Projection of optic flow to normal flow

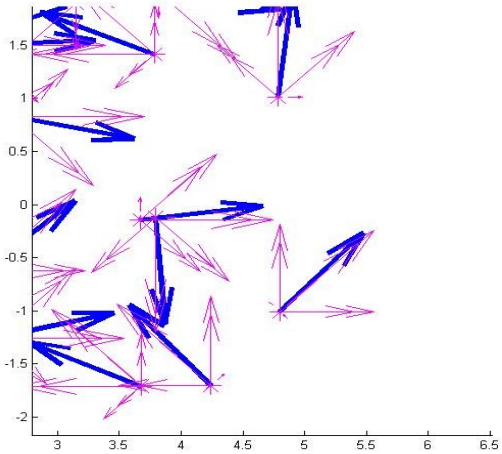
In this example is illustrated the projection of optical flow. The dot is moving is the direction indicated by the **blue** vector (true velocity vector), then are calculated the 8 velocity vectors normal to the local edge of contrast, as previously described. These are represented in **magenta**.

As we can notice, normal vectors differ notably from the original velocity vector.



*An example of the dot implemented in the model: the **blue** bold arrow is the dot's **true velocity vector**, the 8 magenta arrows are the projection of the velocity vector onto the normal vector of the local edge. These **8 new velocity vectors** represent the **normal flow** generated by one dot movement.*

In the following picture are represented both the original optic flow (blue bold arrows) and the calculated normal flow (magenta arrows) for a set of several moving dots in a section 3.5°x3.5° of the aperture.



4.7 Simulation results

Some results of the described algorithm are now shown. On the **left** is reported the optical flow **input** (blue vectors for noise and red vectors for signal) and on the **right** the computation results of **convexity cells** (the brighter, the higher convexity value computed).
From Vaina, Gryzwacz, Saiviroonporn, LeMay, Bienfang, Cowey A. (2003), the paper with patient AMG results, the subject taking the motion discontinuity test has to indicate if the imaginary boundary is present or not. Identification of orientation is not required.
Normal subjects (controls) show that is required at least 10% of signal dots in the display. AMG patient on her right visual field (with lesion) required about 35%.

For all tests, standard parameters are:

- # Dots: **1760**
- # Convexity cells (motion discontinuity detectors) = **5000**
- Threshold: **10% of the maximum convexity value** detected (convexity values below this limit are set to zero)

In the following sections Center radius (**Rc**) and Surround radius (**Rs**) are modulated. In the lower left corner Center and Surround are shown for comparison (yellow Center, blue Surround).

Each set has a specific **percentage of motion net signal** (10%-30%-50%), motion discontinuity orientation is chosen randomly as described in section § 4.5.1.
Color map was normalized for comparison between experiments with/without motion discontinuity.

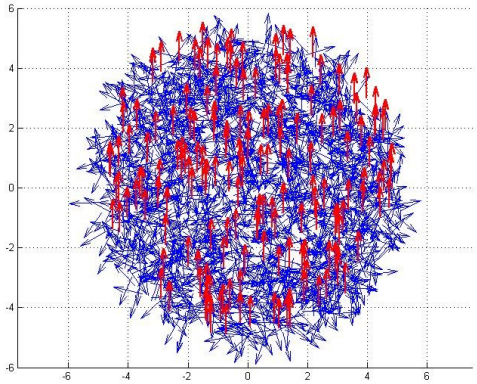
4.7.1 ***Set 1: Motion signal 10%***

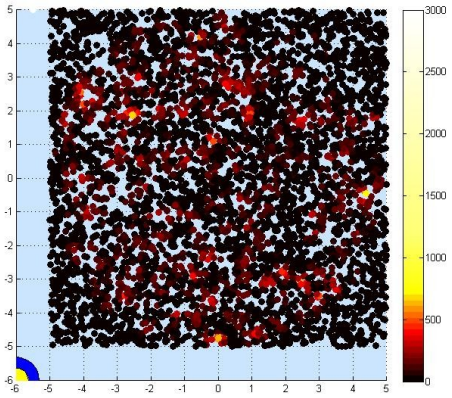
In this set of simulations net signal is 10% of the dots.

4.7.1.1 ***Rc = 0.35° ; Rs = 0.675°***

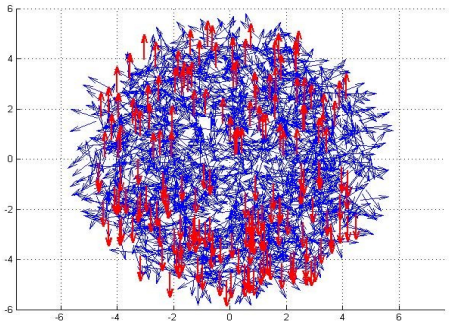
Rc = 0.35° → Center width = 0.70°; Rs = 0.675° → Surround width = 1.35°

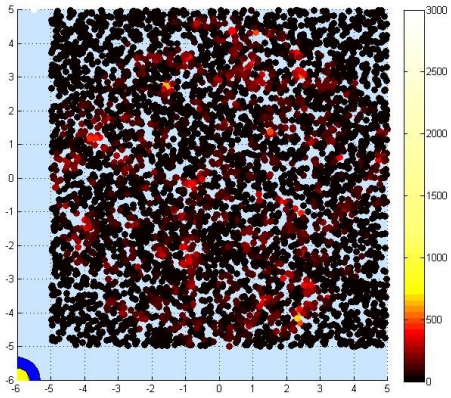
1. no discontinuity





2. horizontal discontinuity

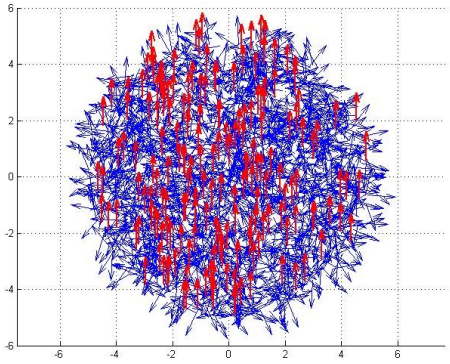


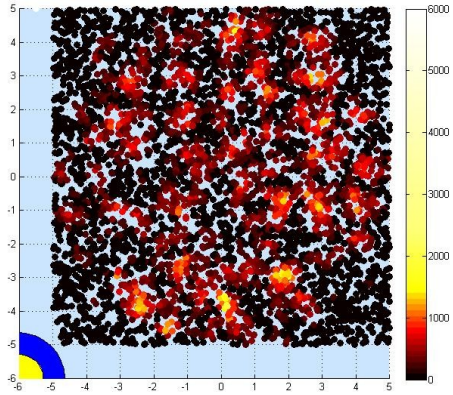


4.7.1.2 ***Rc = 0.70° ; Rs = 1.35°***

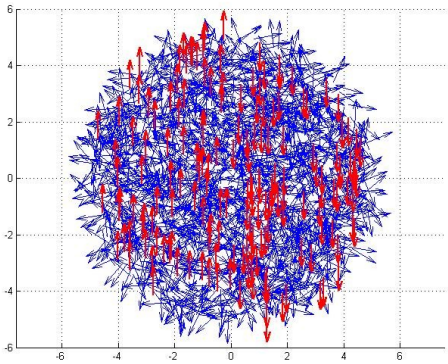
Rc = 0.70° → Center width = 1.40°; Rs = 1.35° → Surround width = 2.70°

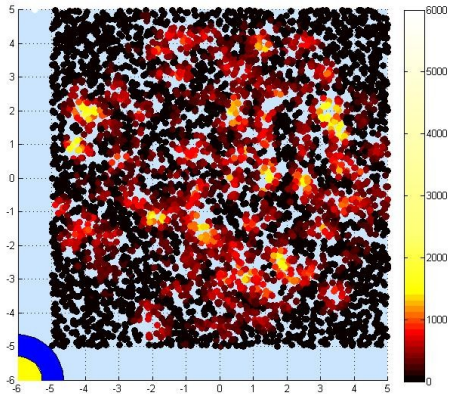
1. no discontinuity





2. vertical discontinuity

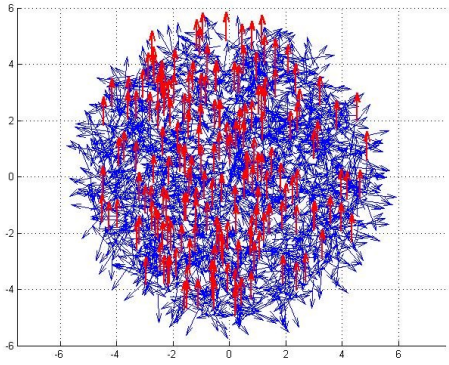


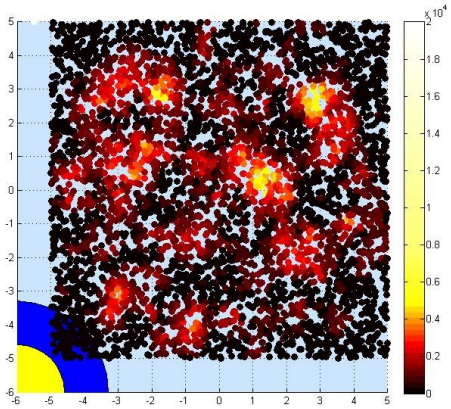


4.7.1.3 ***Rc = 1.40° ; Rs = 2.70°***

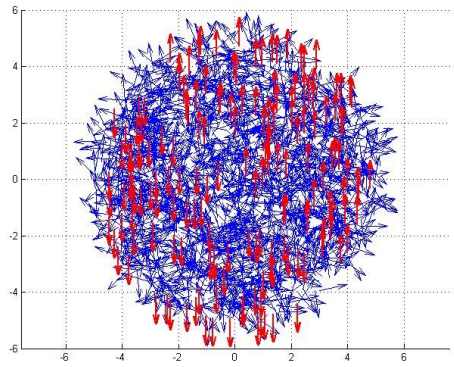
Rc = 1.40° → Center width = 2.80°; Rs = 2.70° → Surround width = 5.40°

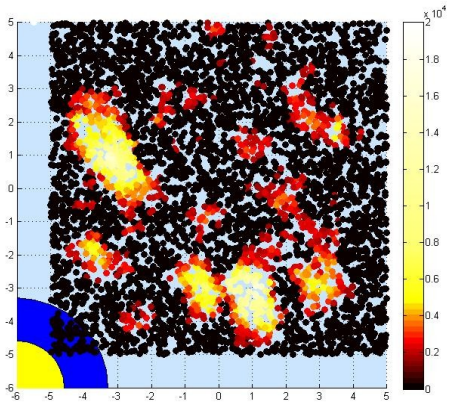
1. no discontinuity





2. -45° discontinuity

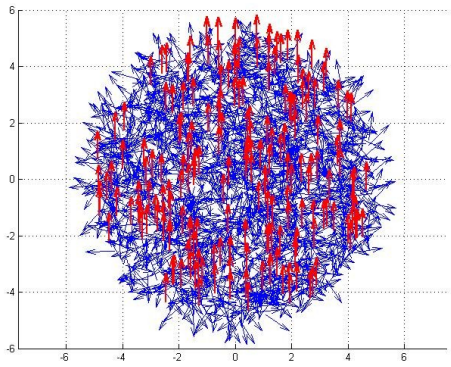


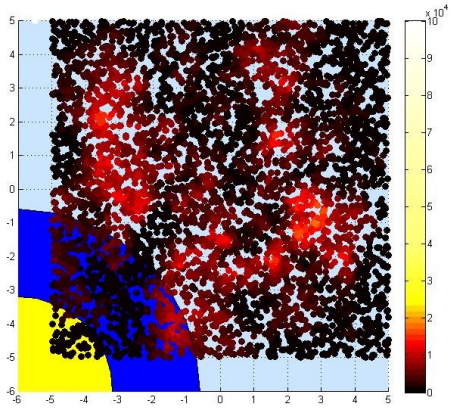


4.7.1.4 ***Rc = 2.80° ; Rs = 5.40°***

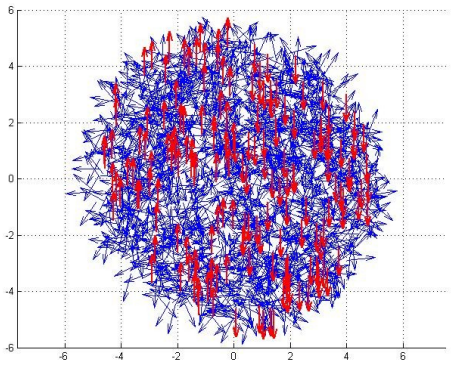
Rc = 2.80° → Center width = 5.60°; Rs = 5.40° → Surround width = 10.80°

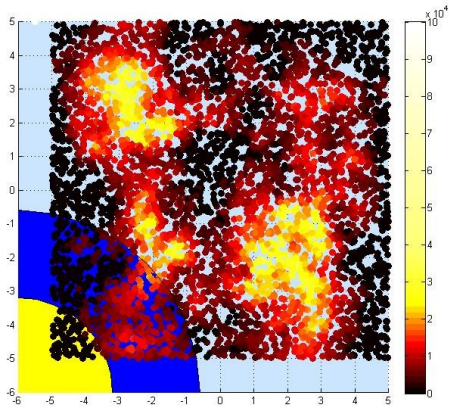
1. no discontinuity





2. vertical discontinuity





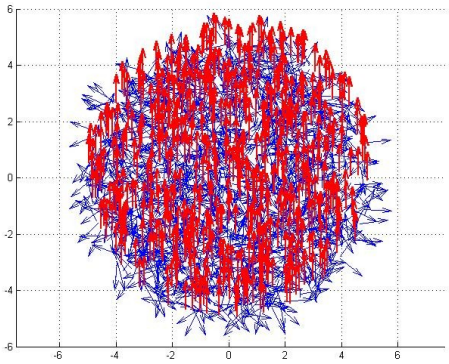
4.7.2 ***Set 2: Motion signal 30%***

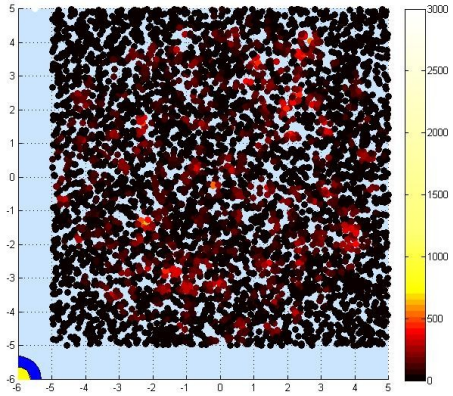
In this set of simulations net signal is 30% of the dots.

4.7.2.1 ***Rc = 0.35° ; Rs = 0.675°***

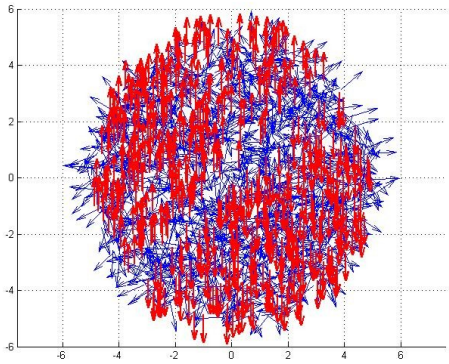
Rc = 0.35° → Center width = 0.70°; Rs = 0.675° → Surround width = 1.35°

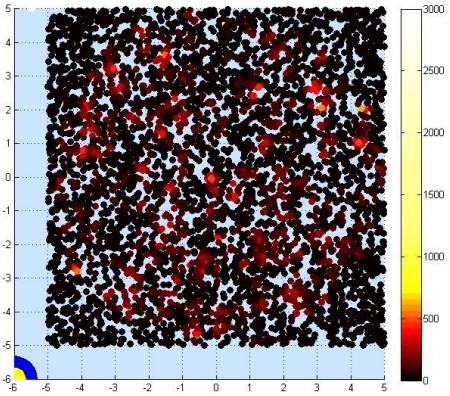
1. no discontinuity





2. 2. +45° discontinuity

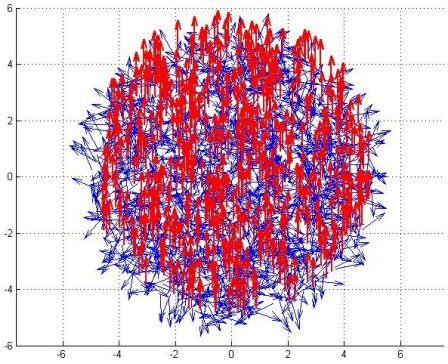


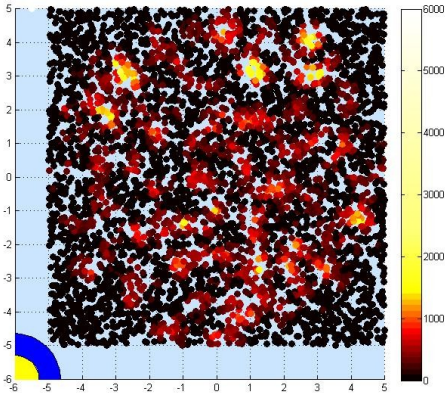


4.7.2.2 ***Rc = 0.70° ; Rs = 1.35°***

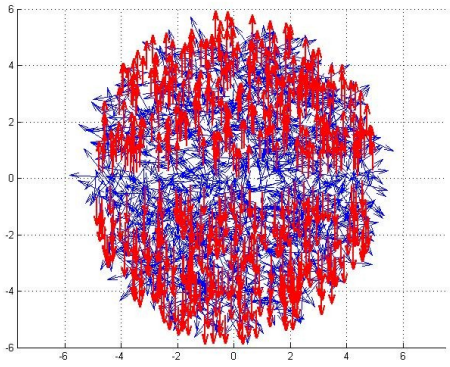
Rc = 0.70° → Center width = 1.40°; Rs = 1.35° → Surround width = 2.70°

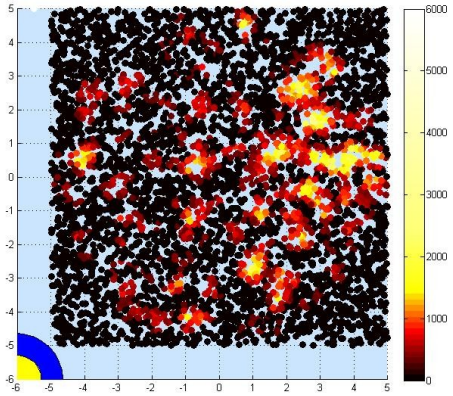
1. no discontinuity





2. horizontal discontinuity

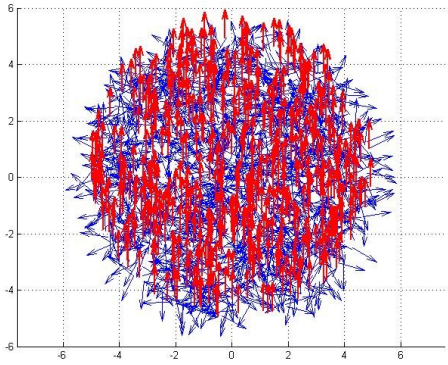


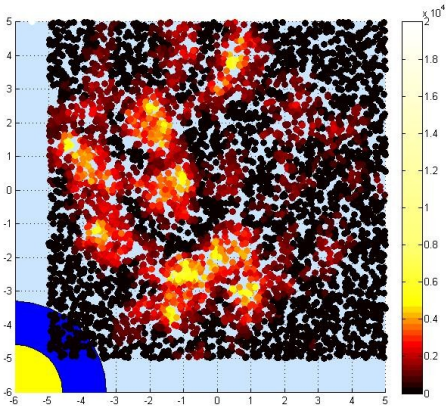


4.7.2.3 ***Rc = 1.40° ; Rs = 2.70°***

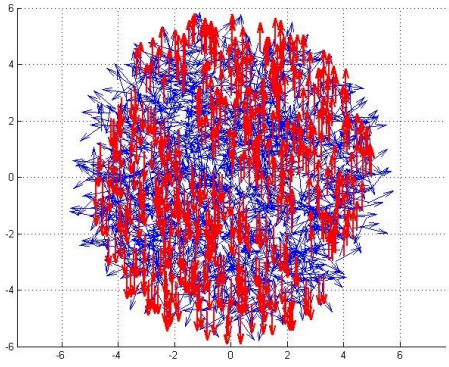
Rc = 1.40° → Center width = 2.80°; Rs = 2.70° → Surround width = 5.40°

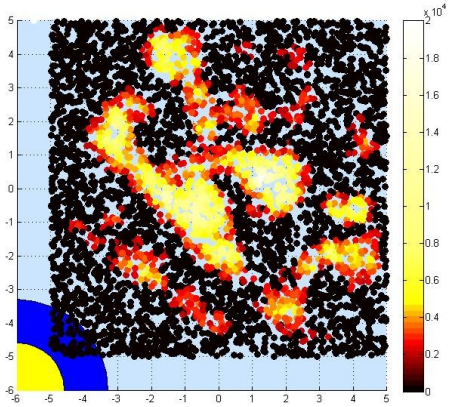
1. no discontinuity





2. -45° discontinuity

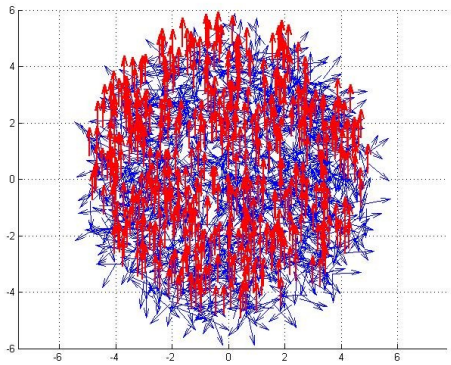


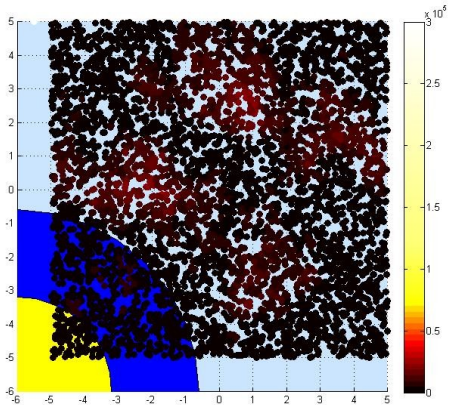


4.7.2.4 ***Rc = 2.80° ; Rs = 5.40°***

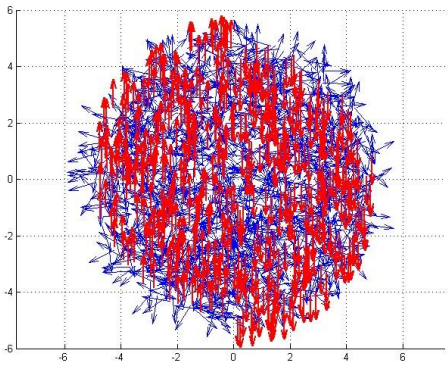
Rc = 2.80° → Center width = 5.60°; Rs = 5.40° → Surround width = 10.80°

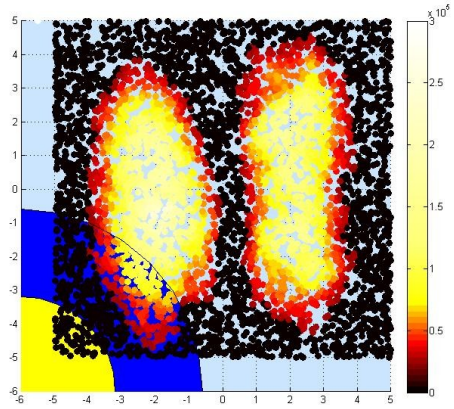
1. no discontinuity





2. vertical discontinuity





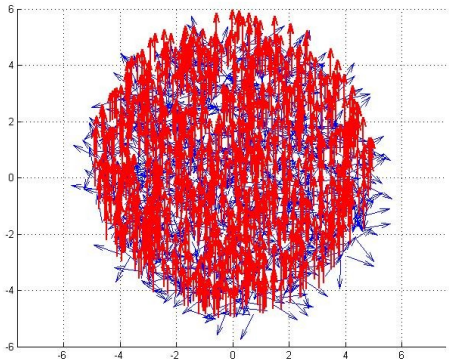
4.7.3 ***Set 3: Motion signal 50%***

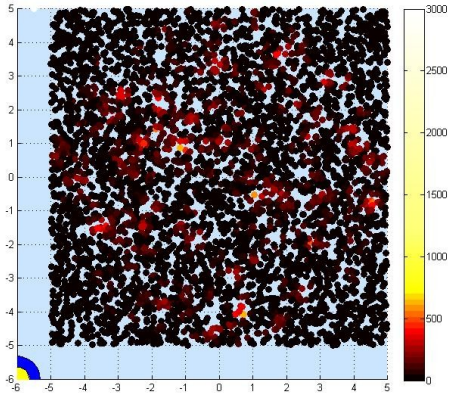
In this set of simulations net signal is 50% of the dots.

4.7.3.1 ***Rc = 0.35° ; Rs = 0.675°***

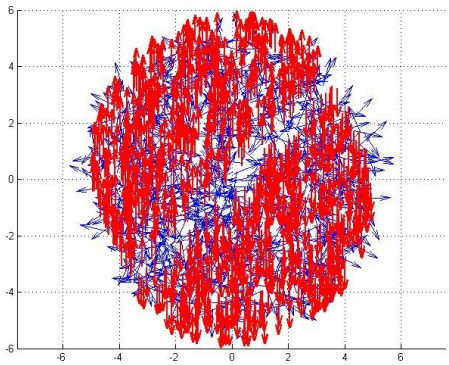
Rc = 0.35° → Center width = 0.70°; Rs = 0.675° → Surround width = 1.35°

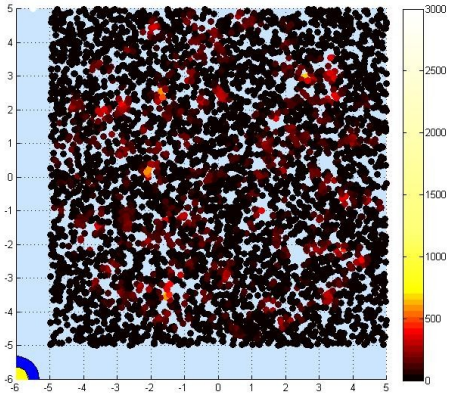
1. no discontinuity





2. +45° discontinuity

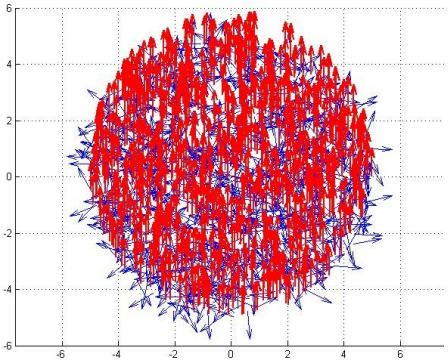


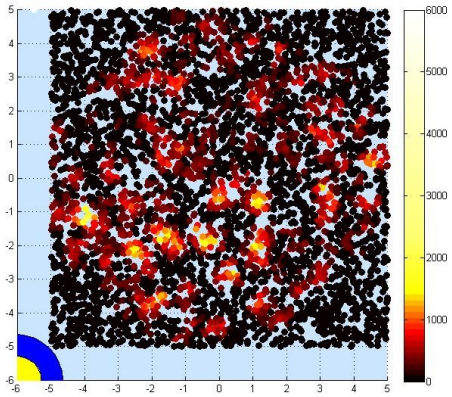


4.7.3.2 ***Rc = 0.70° ; Rs = 1.35°***

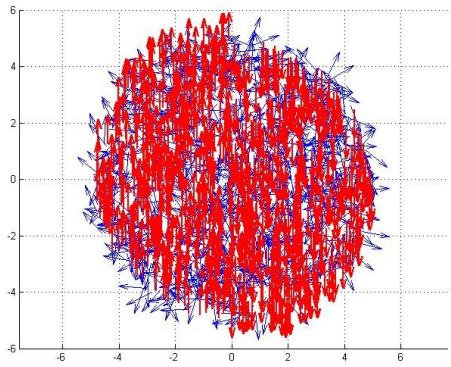
Rc = 0.70° → Center width = 1.40°; Rs = 1.35° → Surround width = 2.70°

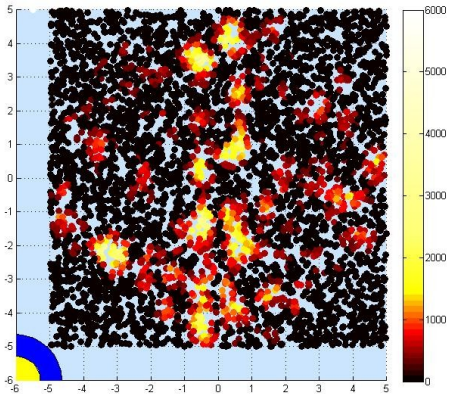
1. no discontinuity





2. vertical discontinuity

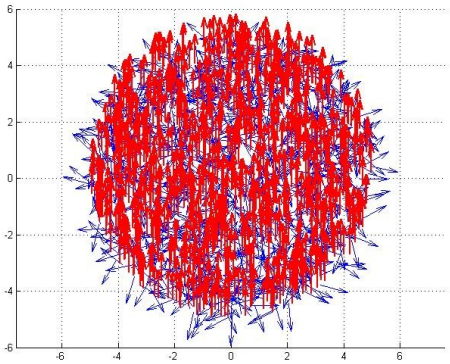


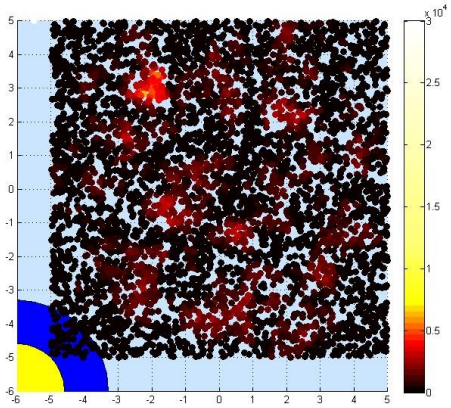


4.7.3.3 ***Rc = 1.40° ; Rs = 2.70°***

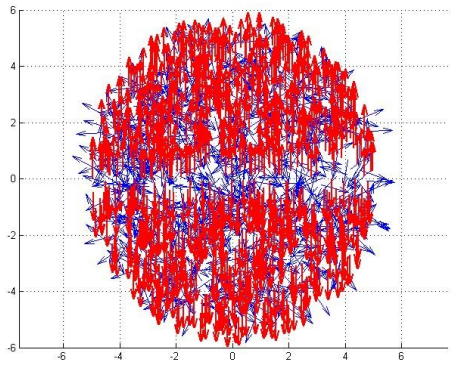
Rc = 1.40° → Center width = 2.80°; Rs = 2.70° → Surround width = 5.40°

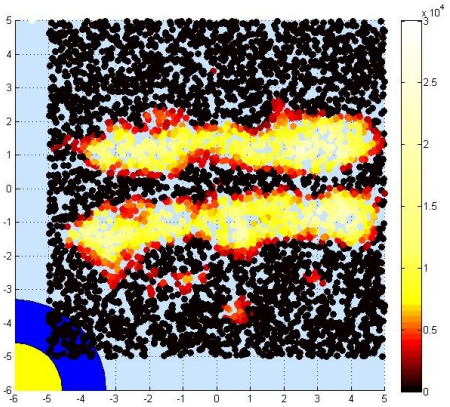
1. no discontinuity





2. horizontal discontinuity

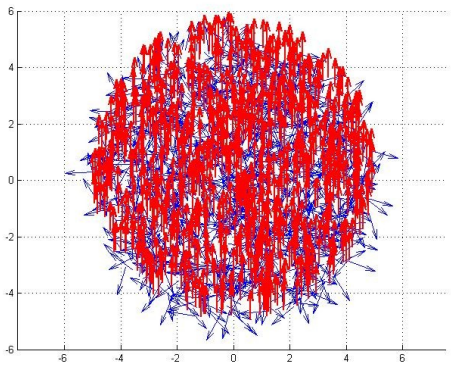


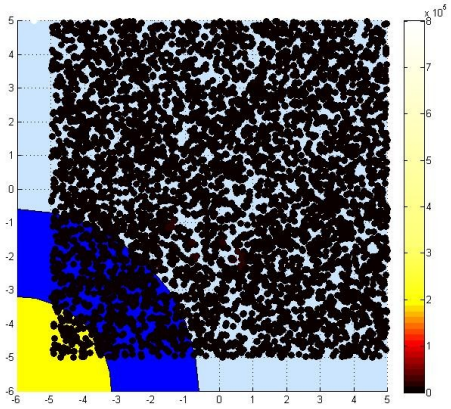


4.7.3.4 *Rc = 2.80° ; Rs = 5.40°*

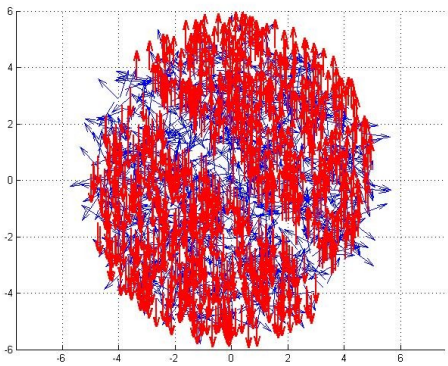
Rc = 2.80° → Center width = 5.60°; Rs = 5.40° → Surround width = 10.80°

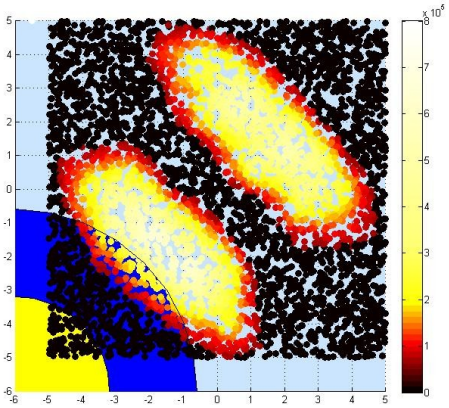
1. no discontinuity





2. -45° discontinuity





4.8 Validation and discussion

For validating the model, further tests are needed to tune the model parameters and to compare them with true patient's data. At present time, no data is available to make any specific comparison.

From the simulations reported earlier we can qualitatively settle some conclusions.

4.8.1 *Importance of the receptive area*

First, Center and Surround receptive field radiuses are extremely important for the performance of the modified Nakayama-Loomis algorithm. Since it is unknown the correct radius, we implemented 4 situations taking as basis the results from Cavanaugh-Bair-

Movshon (2002) studies on Macaque V1 neurons and multiplicating them by 0.50x, 2x and 4x:

1. Very small receptive fields:
   Rc=0.35° (Center width of 0.70°) and Rs=0.675° (Surround width of 1.35°)
2. Cavanaugh-Bair-Movshon small receptive fields:
   Rc=0.70° (Center width of 1.40°) and Rs=1.35° (Surround width of 2.70°)
3. Large receptive fields:
   Rc=1.40° (Center width of 2.80°) and Rs=2.70° (Surround width of 5.40°)
4. Very large receptive fields:
   Rc=2.80° (Center width of 5.60°) and Rs=5.40° (Surround width of 10.80°)

Preliminary, we can see that **very small receptive sizes do not compute motion discontinuity** in any condition.

*4.8.2 **Importance of net motion signal***

Motion discontinuity tests like the one reported in Vaina et al. (2003) show that normal subject in similar conditions of the simulated test, require only 10% of signal dots.
We tested the algorithm in 3 situations:

1. 10% signal – 90% noise
2. 30% signal – 70 % noise
3. 50% signal – 50 % noise

In each situation we tested the 4 types of receptive fields previously described, to check the algorithm behavior at different scales.
In particular, 50% signal is to be considered overabundant even for patients with specific deficits.

*4.8.3 **Discussion of results***

We can state that the algorithm performances increase much more with wide receptive fields than net motion signal. In fact, at **10% signal** using **very large receptive fields** (§4.7.1.4) the results is a clear motion discontinuity detection (high mean convexity value across the aperture) or not detection. In this situation, the algorithm cannot provide an orientation of the discontinuity, but that is consistent with real tests where subject don't have to specify the motion boundary orientation, just it's presence.

Rising the net motion signal to its triple, **30%**, the algorithm still needs **large receptive fields** to clearly detect a motion discontinuity, but it gives a clue of it's orientation too. In this condition, using Cavanaugh-Bair-Movshon Macaque V1 receptive fields lead to very hard-guessing results, with high probability of failure. Using very large receptive fields instead, gives a precise and well defined representation of the motion discontinuity (§ 4.7.2.4). It's interesting to notice that the discontinuity lays in a place of low convexity values (black) and it's surrounded by high convexity values (red-yellow-white). This it's obvious thinking that on the motion boundary the difference $\left[N_i^C - k \cdot N_i^S\right]^2 = 0$ since the exact same optical (and hence normal) flow is on both Center and Surround.

Giving a very strong signal of **50%** does increment the reliably of results using Cavanaugh-Bair-Movshon Macaque V1 receptive fields (§4.7.3.2), though it completely marks the motion boundary using large receptive fields.

Durant-Zanker (2009) found that the best detection of a motion defined Gabor pattern occurs at around 4-5° full width at half height, corresponding to about 8.5° of total receptive field size. This is coherent with our findings, since our best human-comparable results are obtained with a Center width of 5.60° and a Surround width of 10.80°.

These results lead to the conclusion that even though the input, normal flow, comes from local sources, the summation of votes must occur on a quite large scale of about 10°.

Chapter 5 **Open problems and conclusions**

> *"Individual neurons early in the visual system (LGN or V1) respond to motion that occurs locally within their receptive field. Because each local motion-detecting neuron will suffer from the aperture problem, the estimates from many neurons need to be integrated into a global motion estimate. This appears to occur in Area MT/V5 in human visual cortex."* Wikipedia – Motion perception page on March 3rd 2010 (http://en.wikipedia.org/wiki/Motion_perception)

It's commonly assumed by most of the scientific community that global integration of motion clues from V1 neurons is needed in order to have a reliable motion pattern where to compute high order motion perception functions, including motion discontinuity.

Clinical results from post-lesion patients show that this appears to be not true and suggest that motion integration and motion discontinuity detection are decoupled. This thesis discussed and implemented an algorithm that could be biologically computed and that would solve the issue. In our proposal, motion discontinuity is computed prior or parallel to global motion integration (at least at a rough level) and uses only normal flow, the projection of retinal optical flow over the direction normal to the local edge of contrast, that is the only input available at V1 cells level.

In particular, Vaina et al. (1990) studied AF, a 60 y.o. patient with lesions in the temporal-parietal-occipital junction, probably involving MT, who was impaired in motion mechanisms like recognizing a 2x2° figure moving on a dense random pattern, local speed discrimination, motion coherence detection.
Nevertheless, AF could successfully reconstruct a 3D motion from a 2D pattern of moving dots and recognize a human movement from the movement of lights attached to the joints of actors (structure from motion), like the normal controls.
This indicates that **precise** measurement of local motion **is not necessary** for higher-

order structure from motion computations.

Another crucial clinical evidence is Vaina et al. (2003) where is studied patient AMG, 53 y.o., who had a lesion in the left occipital lobe, involving areas V3 and V3A and underlying white matter. She was impaired in her right visual field in local speed discrimination (but not for direction or orientation); determining the number of bumps added to a circumference; determining a 2D form from speed or direction differences; determining the overall direction of motion.
Motion coherence performance was **normal**, but **motion discontinuity** was severely **impaired**, requiring 4 times the percentage of signal dots respect to normal controls.
In particular, about 10% of dots moving coherently are necessary to correctly discriminate the presence of a motion discontinuity in the aperture, while AMG required (for her right visual field) about 40%. This suggests a deficit of integration across spatial scales.
AMG lesion does not include V1 or area MT, possibly areas 18, V2 and V3. fMRI showed that lesion was centered in V3A and V3.
This leads naturally to the conclusion that perhaps motion discontinuities are computed in V3A or V3 and besides that our results indicate that a quite large receptive field is needed in order to detect correctly motion discontinuities. But indicating the cortical areas of motion computation is beyond the objective of this thesis.

In this thesis after a short introduction in Ch.1, we examined the problem of motion discontinuity detection and the aperture problem in Ch. 2. We found that in order to skip the global integration, the only way is using the normal flow as input for our computations.

Then in Ch. 3 we analyzed a series of previous works relevant to the subject, to try to define the actual knowledge of the problem.

Finally in Ch. 4 we illustrated and implemented the algorithm assuming as hypothesis that a biological “Convexity cell” could actually compute it. Some simulation tests were performed and graphical results are shown and discussed. In specific, we found that the algorithm computes correctly motion discontinuities only with large Center and Surround regions, from 5° to 10° wide. This means that if such a cell as a “Convexity cell” existed, its

receptive field should be quite wide, thus realizing not a strictly local computation. On the other side, since it does not require optic flow but only normal flow, it could still be located in a prior and different structure than the one deputed to integration of motion estimates of V1 in order to reconstruct optic flow. This could explain the clinical data of impaired people.

Future studies should create a biologically feasible model, like the one proposed by Beck, Ognibeni, Neumann (2008), recreating the brain structures and implementing the functions that we know of, including the proposed modified Nakayama-Loomis convexity function for the detection of motion boundaries. Then it should be validated with specific tests on normal subjects and fMRI studies.

In the end it would be interesting to introduce into the model irregularities in order to recreate the deficits found in patients. That would be a great indication about the possible cortical functional organization, since we can't study it *in vivo* and the only possible approach is the “black box” with hints given from fMRI.

In conclusion, citing McCool and Britten (2007), “striking is how much remains to be learned” about the motion system in visual cortex. If this thesis could help understanding at least one concept, it would have it's reward.

Chapter 6 **Appendix**

6.1 Apparatus

Toshiba netbook NB100, CPU Intel Atom N270 1.60GHz dual core, 1GB Ram, Windows XP Home SP3.

Rendering of each simulated test (1760 Dots, 5000 Convexity cells) took from 50 sec to more than 120 sec, depending on the center-surround radiuses.

6.2 MatLab code

In this section is reported the full MatLab code developed in this thesis. It should work on any MatLab >6.5 or GNU Octave >3.0

```
% Thesis: A NOVEL METHOD FOR COMPUTING MOTION DISCONTINUITY
% Author: Davide Adamoli under supervision of Prof. Lucia M. Vaina
% Universita' di Padova (Italy) & Boston University (USA)
% April 13th, 2010

clear all;
close all;
% Dot matrix is the datastructure containing the dot's coordinates X and Y,
% magnitude and angle of velocity and if it's a signal dot or not
Dot=[]; %ROW = dot index;
        %COLUMN = property
        %   1=X
        %   2=Y
        %   3=magnitude
        %   4=theta angle
        %   5=isSignal 1=signal, 0=noise
% Edges matrix is the datastructure containing the subdot's coordinates X and Y,
% and the 8 projections of the velocity vector over the 8 possible
% directions considered, due to aperture problem. Clearly, each "Dot"
% produces 8 edges velocities
Edges=[];   %ROW = edge point index;
            %COLUMN = property
            %   1=X
            %   2=Y
            %   3= projection over phi(1)
            %   ...
            %   10=projection over phi(8)
nDots = 1760; %number of total dots on screen
% Number of "Convexity cells" - motion discontinuity detectors
nDetectors = 5000; %total number of Convexity cells in the simulation
r = 5/60; % radius of the dot = 5' = (5/60)°
          % hypothesis is that dots have 10' diameter
```

```
maxMag = 1; % maximum velocity magnitude
aperture = 10; % aperture value in degrees (°)
signalProb = 0.10; % probability for a dot to be a signal dot
                   % 0.50 = 50%
signalMag = 1;              % signal components: mag = 1
signalTheta = deg2rad(90);  % signal direction: upward = 90°
% Convexity variables
Rc=(1.40)^2;  % Center radius in degrees (squared for algorithm optimization)
Rs=(2.70)^2;  % Surround radius
% Set Threshold level: if convexity function is below this limit, is set to
% zero
thresholdlevel=0.1; % 0.1 = 10% of the maximum convexity value
colorcontrol=20000; %mapcolor control (maximum convexity value for white)
% Data structures declaration:
noiseDots=[];
signalDots=[];
%DISCONTINUITY TYPE
discType = 5; %set the type of motion discontinuity
% 1 - HORIZONTAL DISCONTINUITY y=0
% 2 - VERTICAL DISCONTINUITY x=0
% 3 - +45° DISCONTINUITY y=x
% 4 - -45° DISCONTINUITY y=-x
% 5 - NO DISCONTINUITY

nDir = 8; %number of possible directions considered
%phi[] is the array of the directions angles 'phi'
phi(1)=deg2rad(90);
phi(2)=deg2rad(45);
phi(3)=deg2rad(0);
phi(4)=deg2rad(315);
phi(5)=deg2rad(270);
phi(6)=deg2rad(225);
phi(7)=deg2rad(180);
phi(8)=deg2rad(135);

%---------------- CREATING APERTURE AND DOTS --------------------
k=1; % signalDots index
l=1; % noiseDots index
% Create nDots dots with random location and velocity vector signal or noise
for i=1:nDots
    % first of all, the dot location must be inside the 10°x10° aperture
    Dot(i,1)=((aperture)*rand)-aperture/2; % X set random location
    Dot(i,2)=((aperture)*rand)-aperture/2; % Y (uniform distribution)
    while ((Dot(i,1)^2+Dot(i,2)^2)>(aperture/2)^2) %check if it is inside the
circle
            Dot(i,1)=((aperture)*rand)-aperture/2; %if it's not, reassign
location
            Dot(i,2)=((aperture)*rand)-aperture/2; %not very efficient but works
    end

    % now, location is ok, we assign the properties to the new dot
    if (rand<signalProb) %sort if this is a SIGNAL DOT or a NOISE DOT
        Dot(i,5)=1;
    else
        Dot(i,5)=0;
    end
    % set velocity components
    if (Dot(i,5)==1) % CASE of SIGNAL DOT
        switch discType %assign velocity based on the discontinuity type
```

```
            case 1
                if Dot(i,2)>=0 % Y>=0
                    Dot(i,3)=signalMag;
                    Dot(i,4)=signalTheta;
                else
                    Dot(i,3)=signalMag; %identical magnitude
                    Dot(i,4)=signalTheta+(pi); %opposite versus
                end
            case 2
                if Dot(i,1)<=0 % X<=0
                    Dot(i,3)=signalMag;
                    Dot(i,4)=signalTheta;
                else
                    Dot(i,3)=signalMag; %identical magnitude
                    Dot(i,4)=signalTheta+(pi); %opposite versus
                end
            case 3
                if Dot(i,2)>=Dot(i,1) % Y>=X
                    Dot(i,3)=signalMag;
                    Dot(i,4)=signalTheta;
                else
                    Dot(i,3)=signalMag; %identical magnitude
                    Dot(i,4)=signalTheta+(pi); %opposite versus
                end
            case 4
                if Dot(i,2)>=-Dot(i,1) % Y>=-X
                    Dot(i,3)=signalMag;
                    Dot(i,4)=signalTheta;
                else
                    Dot(i,3)=signalMag; %identical magnitude
                    Dot(i,4)=signalTheta+(pi); %opposite versus
                end
            case 5
                % NO DISCONTINUITY, all signal
                    Dot(i,3)=signalMag;
                    Dot(i,4)=signalTheta;

        end
 %signalDots is a support matrix useful for plotting vectors (X,Y,Vx,Vy)
        signalDots(k,1)= Dot(i,1); %X
        signalDots(k,2)= Dot(i,2); %Y
        signalDots(k,3)= Dot(i,3)*cos(Dot(i,4)); %Vx=mag*cos(theta)
        signalDots(k,4)= Dot(i,3)*sin(Dot(i,4)); %Vy=mag*sin(theta)
        % If Vx or Vy are very small, approximate them to zero
        % (due to matlab pi approximation)
        if abs(signalDots(k,3)) < 1.00e-15
            signalDots(k,3)=0;
        end
        if abs(signalDots(k,4)) < 1.00e-15
            signalDots(k,4)=0;
        end

        k=k+1;
    else % CASE of NOISE DOT
%       Dot(i,3)=(maxMag*rand); %random magnitude
        Dot(i,3)= maxMag;       %fixed magnitude
        Dot(i,4)=(2*pi*rand);   %random theta between 0 and 2pi
  %noiseDots is a support matrix useful for plotting vectors (X,Y,Vx,Vy)
        noiseDots(l,1)= Dot(i,1); %X
```

```
        noiseDots(l,2)= Dot(i,2); %Y
        noiseDots(l,3)= Dot(i,3)*cos(Dot(i,4)); %Vx=mag*cos(theta)
        noiseDots(l,4)= Dot(i,3)*sin(Dot(i,4)); %Vy=mag*sin(theta)
        l=l+1;
    end
end
%----------------DOTS CREATED-----------------------

%-----------CREATING EDGE POINTS--------------------
normalFlow=[]; %Normal Flow velocity vectors (X,Y,Vx,Vy)
k=0; %index for normalFlow matrix
for i=1:nDots %i=current Dot
    for j=1:nDir %j=current direction
        %calculate coordinates
        h=(i-1)*nDir+j; %edge point index
        Edges(h,1)=Dot(i,1)+r*cos(phi(j)); %X + radius displacement
        Edges(h,2)=Dot(i,2)+r*sin(phi(j)); %Y + radius displacement

        %calculate projection on directions
        %project only when velocity vector is within 90° of the considered
        %direction, i.e. nu=|phi-theta|, nu<90° or nu>270°
        %else if 90°<nu<270° project onto opposite direction
        %else if nu=90° or nu=270°, set zero (perpendicular)
        nu=abs(phi(j)-Dot(i,4));
        condition1=((nu < deg2rad(90)) | (nu > deg2rad(270)));
        condition2=((nu > deg2rad(90)) & (nu < deg2rad(270)));
        if condition1 %ok, project velocity vector onto direction
            % projection on direction=mag*cos(phi-theta)
            Edges(h,j+2)=Dot(i,3)*cos(nu);
            % this is a new normal flow vector
            k=k+1;
            normalFlow(k,1)=Edges(h,1); %X
            normalFlow(k,2)=Edges(h,2); %Y
            normalFlow(k,3)=Edges(h,j+2)*cos(phi(j)); %Vx
            normalFlow(k,4)=Edges(h,j+2)*sin(phi(j)); %Vy
        elseif condition2 %else project onto opposite direction
            q=(j+(nDir)/2); %index of the opposite direction
            if q>nDir %check if direction is valid
                q=mod(q,nDir);
            end
            Edges(h,q+2)=-Dot(i,3)*cos(nu);
            %this is a new normal flow vector
            k=k+1;
            normalFlow(k,1)=Edges(h,1); %X
            normalFlow(k,2)=Edges(h,2); %Y
            normalFlow(k,3)=Edges(h,q+2)*cos(phi(q)); %Vx
            normalFlow(k,4)=Edges(h,q+2)*sin(phi(q)); %Vy
        else %if none of above, it's perpendicular
            %set zero for this direction
            Edges(h,j+2)=0;
        end
    end
end

%----------------EDGE POINTS CREATED------------------

%_____________________CONVEXITY FUNCTION_______________

% Convexity matrix C(X,Y,Convexity value,... see below)
```

```
% C is the matrix datastructure where all data of convexity cells are
% stored.
k=0;
% regulary spaced Convexity cells
%for i=1:20
%    for j=1:20
%       k=k+1;
%        C(k,1)=-5.5+i*0.5;  %X
%        C(k,2)=-5.5+j*0.5;  %Y
%        C(k,3)=0;  %convexity value init.
%    end
%end
% randomly located Convexity cells
for k=1:nDetectors %k is the number of motion discontinuity detectors
             %and their location is set randomly in the aperture
    C(k,1)= ((aperture)*rand)-aperture/2; %set random location
    C(k,2)= ((aperture)*rand)-aperture/2; %uniform distribution
    C(k,3)=0;  %convexity value init. to zero
end


%Calculate CONVEXITY FUNCTION for all locations in Convexity matrix
%this operation can be done in parallel since each motion discontinuity
%detector is independent from the others. Here we calculate convexity
%value for each detector.
[rowsC colsC] = size(C); %rowsC=total number of convexity cells
%-begin mark A
for h=1:rowsC %repeat for all convexity cells
    alfa=C(h,1);    %location X of the convexity cell
    beta=C(h,2);    %location Y of the convexity cell

% Nc and Ns are arrays where we store the votes took from every direction
% in Center (Nc) and Surround (Ns)
%Nc=zeros(1,nDir);  %Center array init. to 0
%Ns=zeros(1,nDir);  %Surround array init. to 0
Nc=ones(1,nDir);  %Center array init. to 1
Ns=ones(1,nDir);  %Surround array init. to 1
[rowsE colsE] = size(Edges); %rowsE=total number of edge points
for i=1:rowsE %repeat for all edge points
    %distance^2=(X-alfa)^2 + (Y-beta)^2
    dSquared=(Edges(i,1)-alfa)^2 + (Edges(i,2)-beta)^2;
% evaluate if this edge point is in Center or Surround of this convexity
% cell. If so, count its votes in Nc or Ns arrays.
    if dSquared<=Rc    %Edge point in Center region
        for j=1:nDir
            if Edges(i,j+2) > 0.0001 %j-th direction gets a vote
                Nc(j)=Nc(j)+1;
            end
        end
    elseif dSquared<=Rs    %Edge point in Surround region
        for j=1:nDir
            if Edges(i,j+2) > 0.0001 %j-th direction gets a vote
                Ns(j)=Ns(j)+1;
            end
        end
    end %if this Edge point is neither Center or Surround, do nothing
end

%in Nc and Ns are the votes for (alfa, beta) location,
```

```
%Nc(j)= total n. of votes in the j-th direction in Center region
%Ns(j)= total n. of votes in the j-th direction in Surrounding region
%calculate Convexity function
%C(alfa,beta)=SUMi(Nci - k*Nsi)^2

% CALCULATE k=SUMi Nc(i) / SUMi Ns(i)
% k takes into account the absolute difference of votes between Center and
% Surround
sumC=sum(Nc);
sumS=sum(Ns);
k=sumC/sumS;

for j=1:nDir % repeat for each direction
    %k=(Nc(j)/sum(Nc))/(Ns(j)/sum(Ns)); %k specific for this direction
    C(h,3)=C(h,3)+((Nc(j)-k*Ns(j))^2); % SQUARED DIFFERENCE
    % C(h,3)=C(h,3)+abs(Nc(j)-k*Ns(j)); % ABSOLUTE DIFFERENCE
    C(h,3+j)=Nc(j); %report the Center votes - test
    C(h,3+9)=8; % marker for test
    C(h,3+9+j)=Ns(j); %report the Surround votes - test
    C(h,3+9+9)=8; % marker for test
    C(h,3+9+10)=sumC; % report the total votes in Center - test
    C(h,3+9+11)=sumS; % report the total votes in Surround - test
    C(h,3+9+12)=k; % report k ratio between Center/Surround
end
%C(h,1) = X location (alfa)
%C(h,2) = Y location (beta)
%C(h,3) = convexity vaue for location (alfa,beta)

end %-end mark A
%_________________END OF CONVEXITY FUNCTION_________________


%---------------PLOTTING RESULTS----------------------
S=0; %scale for plotting vectors
s=0;
scrsz = get(0,'ScreenSize');

% first figure: normal flow
fig1=figure('Position',[1 scrsz(4)/2 scrsz(3)/2 scrsz(4)/2]);
for i=1:nDots
    DotVx(i)=Dot(i,3)*cos(Dot(i,4));
    DotVy(i)=Dot(i,3)*sin(Dot(i,4));
    %If Vx or Vy are very small, approximate them to zero
    %(due to matlab pi approximation)
    if abs(DotVx(i)) < 1.00e-15
        DotVx(i)=0;
    end
    if abs(DotVy(i)) < 1.00e-15
        DotVy(i)=0;
    end
end
hold on;

%plot true optical flow
%quiver(Dot(:,1), Dot(:,2), DotVx, DotVy, S, 'b', 'LineWidth', 4);

%plot normal flow
quiver(normalFlow(:,1),normalFlow(:,2),normalFlow(:,3),normalFlow(:,4), s,'m');
hold off;
```

```
axis equal;

%second figure: plot noise and signal Dots with respective velocities
fig2=figure('Position',[scrsz(3)/2 scrsz(4)/2 scrsz(3)/2 scrsz(4)/2]);
hold on;
grid on;
axis([-6 6 -6 6]);
axis square;
quiver(noiseDots(:,1), noiseDots(:,2), noiseDots(:,3), noiseDots(:,4), S,'b' );
if length(signalDots) ~= 0 %plot only if there's any signal dot to plot
quiver(signalDots(:,1), signalDots(:,2), signalDots(:,3), signalDots(:,4), S,
'r', 'LineWidth', 2 );
end
hold off;
axis equal;

%third figure: plot computed CONVEXITY VALUES
fig3=figure('Position',[1 1 scrsz(3)/2 scrsz(4)/2]);
hold on;
%A(X,Y,Convexity value) is a support matrix for plotting X and Y
A=[C(:,1),C(:,2)];
maxConv=max(C(:,3)); %maximum convexity value computed

threshold=thresholdlevel*maxConv; %set threshold

[rowsC colsC] = size(C);
for h=1:rowsC
%THRESHOLD convexity values
    if C(h,3) < threshold
        A(h,3)=0; %if value is under a certain threshold, set to zero (black, no
signal)
    else
        A(h,3)=C(h,3); %else, set a convexity signal (firing rate?)
    end
end
% colormap control dot for adjusting colorspace
A(rowsC+1,1)= -5.5; %x
A(rowsC+1,2)= 5; %y
%MAPCOLOR CONTROL
A(rowsC+1,3)= colorcontrol; %max color value for colormap max value

%plot3(A(:,1),A(:,2),A(:,3));
%stem3(A(:,1),A(:,2),A(:,3));
%load seamount;
scatter(A(:,1),A(:,2),40,A(:,3), 'filled');
%contour(C(:,1),C(:,2),C(:,3));
%set(h,'ShowText','on','TextStep',get(h,'LevelStep')*2)
%colormap cool
%Env=
%image(Env,'CDataMapping','scaled')
%colormap(gray)
%axis image

% Plot 2 circles corresponding to the Center and Surround areas for
% comparison
t = (0:1/32:1)'*2*pi;
%Surround: BLUE
x = sqrt(Rs)*sin(t)-6;
y = sqrt(Rs)*cos(t)-6;
```

```
fill(x,y,'b');
%Center: YELLOW
x = sqrt(Rc)*sin(t)-6;
y = sqrt(Rc)*cos(t)-6;
fill(x,y,'y');


hold off;
grid on;
axis([-6 5 -6 5]);
axis square;
load('DadeMap2','cm2'); % use a special colormap
set(fig3,'Colormap',cm2);
% cm associated with DadeMap
% cm2 associated with DadeMap2 - has more color points
%cm=colormap(hot);
colorbar;
set(gca,'Color',[0.8 0.9 1]); %set background color
%axis equal;
```

Chapter 7 **References**

Beardsley, S.A.; Vaina, L.M. (2001) “A Laterally Interconnected Neural Architecture in MST Accounts for Psychophysical Discrimination of Complex Motion Patterns”, Journal of Computational Neuroscience 10, 255-280, 2001

Beck, C.; Ognibeni, T.; Neumann, H. (2008) “Object Segmentation from Motion Discontinuities and Temporal Occlusions – A Biologically Inspired Model”, PloS ONE 3(11):e3807

Cavanaugh, J.R.; Bair, W.; Movshon, J.A. (2002) "Nature and Interaction of Signals From the Receptive Field Center and Surround in Macaque V1 Neurons", Journal of Neurophysiology 88:2530-2546, 2002

Clifford, C.W.G.; Ibbotson, M.R. (2003) “Fundamental mechanisms of visual motion detection: models, cells and functions”, Progress in Neurobiology 68 (2003) 409-437

Durant, S.; Zanker, J.M. (2009) “Characterizing motion contour detection mechanisms and equivalent mechanisms in the luminance domain”, Journal of Vision 9(1):36

Ganong, W. F. (2006) “Fisiologia Medica”, 10th Italian ed., Piccin Nuova Libraria S.p.A. Padova, based on “Review of Medical Physiology”, 21st ed.

Gibson, J. (1966) “The senses considered as perceptual systems”, Houghton Mifflin, Boston

Grzywacz, N.M.; Yuille, A.L. (1990) “A model for the estimate of local image velocity by cells in the visual cortex”, Proceedings of the Royal Society of London. Series B, Biological Sciences, vol. 239, No. 1295, 129-161

Guyton, A.C.; Hall, J.E. (2002) “Fisiologia Medica”, 2nd Italian ed., EdiSES s.r.l. Napoli, based on “Textbook of Medical Physiology”, 10th ed.

Hildreth, E. C. (1983) “The Measurement of Visual Motion”, The MIT Press

Koch, C.; Wang, H.T.; Mathur, B. (1989) “Computing motion in the primate's visual system”, J. exp. Biol. 146, 115-139 (1989), The Company of Biologists Limited

Logothetis, N.K. (1999) “Vision: a window on consciousness”, Scientific American November 1999

Marr, D.; Poggio, T.; Hildreth, E.C. (1980) “Smallest Channel in Early Human Vision”, Journal of the Optical Society of America 70, 868-870

Marr, D.; Ullman, S. (1981) “Directional Selectivity and its Use in Early Visual Processing”, Proceedings of the Royal Society of London B. 211, 151-180

Majaj, N.J.; Carandini, M.; Movshon, J.A. (2007) “Motion Integration by Neurons in Macaque MT Is Local, Not Global”, The Journal of Neuroscience, January 10, 2007 – 27(2):366-370

McCool, C.H.; Britten, K.H. (2008) “Cortical Processing of Visual Motion”, “The senses: a comprehensive reference”, Vision II, Vol. 2 (2008), pp. 157-187. Ed. Elsevier Inc.

Nakayama, K.; Loomis, J.M. (1974) “Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis”, Perception 1974, volume 3, pages 63-80

Newsome, W.T.; Paré, E.B. (1988) “A selective Impairment of Motion Perception Following Lesions of Middle Temporal Visual Area (MT)”, The Journal of Neuroscience, June 1988, 8(6): 2201-2211

Poggio, T. (1983) “Visual Algorithms” in: Physical and Biological Processing of images, ed. Braddick and Sleigh, Springer-Verlag, Berlin

Rust, N.C.; Mante, V.; Simoncelli, E.P.; Movshov, J.A. (2006) “How MT cells analyze the motion of visual patterns”, Nature Neuroscience vol.9, number 11, November 2006

Simoncelli, E.P.; Heeger, D.J. (1998) “A Model of Neuronal Responses in Visual Area MT”, Vision Research, 38(5): 743-761, 1998

Spoerri, A. (1991) “The Early Detection of Motion Boundaries”, MIT Artificial Intelligence Laboratory – Technical Report 1275

Taylor, W.R.; Vaney, D.I. (2003) “New directions in retinal research”, Trends in Neurosciences Vol. 26 7 July 2003, Elsevier

Vaina, L.M.; LeMay, M.; Bienfang, D.C.; Choy, A.Y.; Nakayama, K. (1990) “Intact “biological motion” and “structure from motion” perception in a patient with impaired motion mechanisms: A case study”, Visual Neuroscience (1990), 5, 353-369

Vaina, L.M.; Gryzwacz, N. M.; LeMay, M.; Bienfang, D.; Wolpow, E.; (1998) “Perception of motion discontinuities in patients with selective motion deficits”, Watanabe Takeo (Ed.), High-level motion processing: computational, neurobiological, and psychophysical perspectives, MIT Press, pp. 213-247

Vaina, L.M.; Gryzwacz, N. M.; Saiviroonporn, P.; LeMay, M.; Bienfang, D.C.; Cowey A. (2003) “Can spatial and temporal motion integration compensate for deficits in local motion mechanisms?”, Neuropsychologia 41 (2003) 1817-1836

Wallish, P.; Movshon, J.A. “Structure and Function Come Unglued in the Visual Cortex”, Neuron 60, October 23, 2008

Thanks to...

Prof. Lucia Vaina, for her support and for giving me the opportunity to write a thesis in the fascinating area of motion perception despite my chronic lack of time and long periods without reply.

Prof. Alfredo Ruggeri, for his encouraging and support along all my experience at Boston University, from the presentation letter to its conclusion with this thesis.

Un ringraziamento particolare va alla mia famiglia che mi ha supportato/sopportato in questi -troppi- anni universitari. Ho fatto di testa mia, ma alla fine penso sia andata bene così, nonostante tutto.

A Francesca, incontrata proprio tra i banchi universitari, che mi ha fatto crescere, mi ha spronato con amore e grinta e mi ha reso migliore. E che continuerà a farlo.

A Fox e Mario, per l'amicizia vera, di quelle che non si perdono con gli anni, con la distanza o quando tutto cambia, perché in fondo non cambia nulla.

Ai colleghi del Comune di Arzignano, quasi una seconda famiglia che mi ha accolto con entusiasmo e che ogni giorno mi ricorda quanto sia importante l'armonia sul luogo di lavoro.

Ai colleghi dell'ULSS 6 Vicenza, per avermi fatto capire cosa significa lavorare nella Sanità e per avermi tutti insegnato molto, nessuno escluso.

All'Associazione Italiana Arbitri, altra passione incontrata negli anni universitari, per gli insegnamenti di vita, le amicizie e le soddisfazioni regalate.

Sono moltissime le persone che ho incontrato in questa esperienza universitaria. Da tutte ho ricevuto qualcosa, nei più svariati ambiti della vita. A conclusione di questa tesi vorrei citarli tutti, ma temo di perderne qualcuno per strada. Mi limito a dire sono grato a coloro che mi sono stati e mi stanno vicini, perché la formazione di una persona non finisce con una laurea, prosegue nelle relazioni che si instaurano e le condivisioni che si trasmettono.

This thesis was entirely composed using open source software. Gimp was used for graphics, OpenOffice 3 was used for writing, composing and presenting the thesis.

Contacts: *davide@adamoli.net ; vaina@bu.edu*