3

Bio-inspired Connectionist Modelling: An Application to Visual Perception of Motion

Claudio Castellanos Sánchez and Pedro Luis Sánchez Orellana
Laboratory of Information Technologies (LIT) of Cinvestav - Tamaulipas
México

1. Introduction

The visual system of human beings has been optimised through millions of years by natural selection. This helps us to detect the pattern of 3D moving objects, its depth, speed and direction estimation, etc. The research in connectionism is inspired by complexity of neural interactions and their organisation in the brain that can allow us to propose a feasible neuromimetic model to imitate the capacities of human brain.

Although visual perception of motion has been an active research field for the scientific community (since motion is fundamental for most machine perception tasks) [McCane, 2001], recent research on computational neuroscience has provided an improved understanding of human brain functionality. In the human brain, the motion is perceived as an interaction between several cortical areas and in two main pathways : the dorsal pathway formed by primary visual area (V1), middel temporal area (MT), middel superior area (MST), etc., specialized on the detection of motion. The ventral pathway, formed by primary visual area (V1), secondary visual area (V2), third visual area (V3), inferotemporal area (IT), etc., which processes characteristics related to the form of the visual information.

This visual information has been taken to create the so called bio-inspired algorithms, which are based on or inspired by functions of some areas in the brain. This bio-inspired algorithms have been proposed to mimic the abilities of the brain for motion perception and understanding [Castellanos-Sánchez, 2005]. There are several bio-inspired models for visual perception of motion, some of them inspired by V1 neurons with a strong neural cooperative-competitive interactions that converge to a local, distributed and oriented auto-organisation [Fellez & Taylor, 2002; Moga, 2000]. Some others are inspired by MT neurons with cooperative-competitive interactions between V1 and MT and an influence range [Derrington & Henning, 1993; Mingolla, 2003]. And the others are inspired by MST for coherent motion and egomotion detection [Pack et al., 2000; Zemel & Sejnowski, 1998], see [Castellanos-Sánchez, 2005] for a more detailed explanation.

All these works are based on a specific cortical area. However, for our proposal we considered that all these specializations might be integrated to make a more robust process. Here we present a bio-inspired connectionist approach, called CONEPVIM model (Neuromimetical Coneccionist Model for the Visual Perception of Motion), which not only considers the higher areas of processing, but also the information from V1, since it might add descriptors for the motion detection problem, based on a particular adaptation of the

spatiotemporal Gabor filters. Also it takes advantage of a modular and strongly localized approach for the visual perception of motion that handles a shunting inhibition mechanism (based on MT and MTS). Due to the methodology used in the model three neuromimetic indicators emerged for visual perception of motion (proposed [Castellanos-Sánchez, 2005; Castellanos-Sánchez et al., 2004]), they allowed us to identify: null motion in objects, whether the motion in the scene is caused by moving objects or ego-motion, and also the speed and direction of the motion.

In this chapter we discuss about some neurobiological principles, next we mention the foundations for the CONEPVIM model, then we continue with the manipulation of several parameters obtained in the antagonist interaction mechanism and three neuromimetic indicators for motion estimation are shown, these indicators emerge from the interactions between the neurons of V1, MT and MST mainly. A series of experiments with real image sequences are described. Finally, we make some conclusions about the proposed methodology and the results.

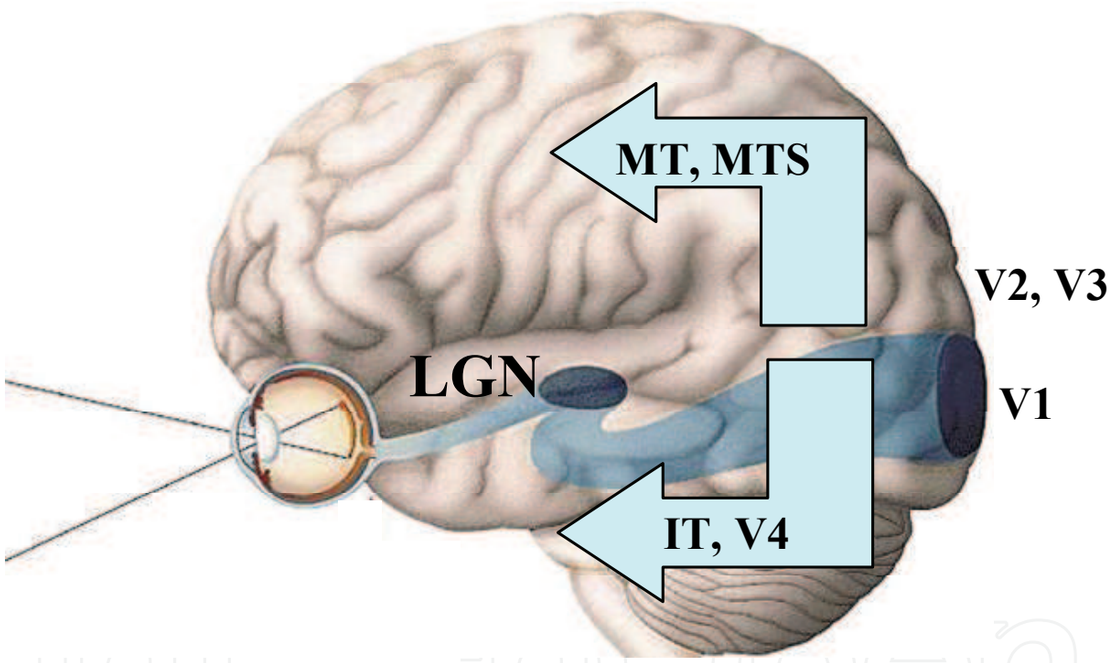


Figure 1. The visual pathway. The information comes in from the retina, after it is integrated in LGN, treated in V1, and finally processed in two different pathways : dorsal (MT, MTS, etc.) and ventral (V4, IT, etc.)

2. Biological foundations

In order to understand better how the bio-inspired model that we propose as well as the existing model work it is necessary to describe some biological bases that sustain them. We separated these foundations in two subsections mentioned in the following: the course of the light signals in the human brain, and the bio-inspiration modelling from the visual pathway.

2.1 The course of the light signals in the human brain

From the retina up to the cerebral cortex of a human being, seventeen different areas take part in vision processing [Sunaert, 1999]. The main areas may be organized in four major stages: acquisition and compression of light signals in the retina; their relaying in the lateral geniculate nucleus (LGN); their cortical analysis in V1, and their secondary treatment in areas MT and MTS of the temporal cerebral cortex in the areas IT and V4 of the parietal cerebral cortex (see figure 1).

a) Acquisition and compression of the light signals in the retina.

In the eye the information from the world is received by photoreceptors, they are called the cones and rods cells. It is from this acquisition that the integration starts by means of the interactions between horizontal and bipolar cells. Finally the information is compressed in a proportion of 160:1 in the ganglional cells [Imbert, 19883].

b) Integration in LGN.

The information comes out from the retina through the optical nerve (formed by the ganglional cells) and receives the next treatments:

- A binocular selection at a optical chiasm, it shares partial information from both eyes.
- A temporal integration in the LGN, which acts as a relay.

80% of the information received in LGN comes from the feedback from the higher areas, and only 20% arrives from the retina [Castellanos-Sánchez, 2005].

c) Analyse in V1.

After the relay in LGN the information of luminal signals is concentreted in the primary visual area (V1), and receives its first cortical treatment by means of the cells sensitive to the contrast (simple cells). These cells have been modelled by local motion energy filters (Gabor filters) [Adelson & Bergen, 1985; Heeger, 1987].

d) Secondary treatment.

Here two pathways may be discovered in this course of visual signals [Hengyi, 2003]: the ventral or occipito-temporal pathway and the dorsal for occipito-parietal pathway. The first one mostly consists of parvocellular cells and it is responsible for the perception of the objects and of their shape, and the second one mostly consists of magnocellular cells and it is responsible for motion and space perception. In the present study, we are particularly interested in the dorsal pathway and in areas specialized in motion analysis.

The processing of the visual signals in the brain might be summariezed in the figure 2.

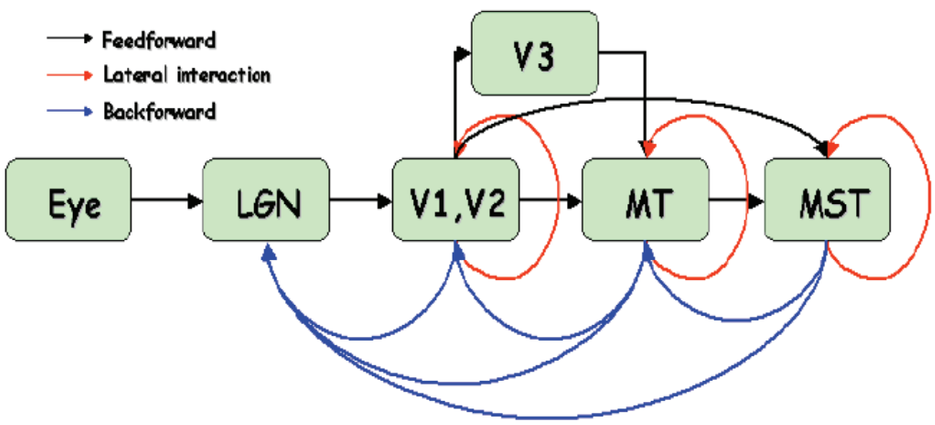


Figure 2. Simplified description of the visual pathway for motion treatment, starting from the eye and ending with the processing in the MST

2.2 Bio-inspired modelling from vision

An important fact is that all the processes mentioned before might be integrated to comform the so called bio-inspired computational models for motion detection, by using local detections and integrating various directions for various scales and spaces to end in a global answer [Van Santen, 1985; Adelson, 1985]. Motion detection, spatiotemporal local inhibition, and integration are the main ideas of neurosciences research that will inspire our connectionist conception. From a biological point of view it is known that motion detection and analysis are achieved by means of a cascade of neural operations [Sekuler, 2002], called: the detection of local motion signals within restricted regions of the visual field and their integration into more global descriptions of the direction and speed of object motion.

The main areas that we are considering in this paper are the human brain areas specialized in motion perception are [Sekuler, 2002]: V1, MT, MST, the kinetic occipital area (KO), and finally the superior temporal sulcus (STS).

The first cortical analysis is performed in V1 by ensuring contrast sensitivity thanks to extended receptive fields. These neurons mainly send their extensions in the vertical direction and they are tuned to a preferred direction of motion [Hubel, 1962] so they perform a local analysis of motion energy that is called filtering. These orientation-selective cells may be modelled as spatiotemporal filters [Adelson, 1985; van Santen, 1984] and their receptive fields may be modelled as a product of inhibitory and excitatory interactions in space and time. On the other hand, contrast detection is sufficient for the identification of motion direction, so that the visual mechanisms that extract motion are built from direction-selective primitives [Derrington, 1993].

Summarizing, the process starts from the local motion of a retinal image that is extracted by neurons in V1 that have a receptive field similar to a small spatially bounded window where they can detect the presence of movement in a specific direction. This strongly localized processing based on lateral interactions is our first source of inspiration for motion detection and estimation. However, the visual perception of motion is not completely determined by the local responses in the neural receptive fields. These responses are also handled to obtain speed information after having collected and combined them from V1 and after having grouped them together in MT. The ambiguity of individual neural responses is solved by this combination of signals.

In the next section we make wider description of the functioning of the stages in the CONEPVIM model, starting from the Casual spatial-temporal Gabor-like filters to finish with the Antagonist interaction mechanism.

3. CONEPVIM model

This section broadly describes the mathematical and biological foundations of the proposed bio-inspired model for visual perception of motion based on the neuromimetic connectionist model reported in [Castellanos-Sánchez, 2005; Castellanos-Sánchez, 2004]. The first stage of this neuromimetic model is mainly based on the causal spatiotemporal Gabor-like filtering and the second stage is a local and massively distributed processing defined in [Castellanos-Sánchez, 2004], where they have proposed a retinotopically organised model of the following perception principle: local motion information of a retinal image is extracted by neurons in the primary visual cortex, V1, with local receptive fields restricted to small areas of spatial interactions (first stage: causal spatio-temporal filtering, CSTF); these neurons are

densely interconnected for excitatory-inhibitory interactions (second stage: antagonist interaction mechanism, AIM).
We will describe in this section the stages of the CONEPVIM model: the spatial processing to model the orientation-selective neurons of V1, temporal processing to model the speed selectiveness of neurons in the medium temporal area, MT, connectionist processing to mimic the excitatory-inhibitory local interactions in the cerebral cortex of human beings, and self-organising mechanisms for coherent motion estimation.

3.1 Casual spatial-temporal Gabor-like filtering (CSTF)

Receptive fields modeled by bidimensional spatial Gabor filters were proposed by Marcelja [Marcelja, 1982] and they were discovered in biological vision systems [Pollen, 1981]. The spatial part of a standard Gabor function approximates well the spatial profile of receptive fields in the cerebral cortex [Jones, 1987]. But its temporal part is non-causal (negative weights are assigned to immediate past images). Several more causal approaches have been proposed. Adelson and Bergen used an asymmetric spatial distribution [Adelson, 1985] and Gzrywacz and Yuille made it with Gabor functions in spatiotemporal co-dependence [Grzywacz, 1990; Grzywacz, 1991]. They concluded that directional selectiveness is equal to orientation in space-time. Our approach handles causality in a simple and local way with a strong hypothesis that ensures the ability to detect local motions.

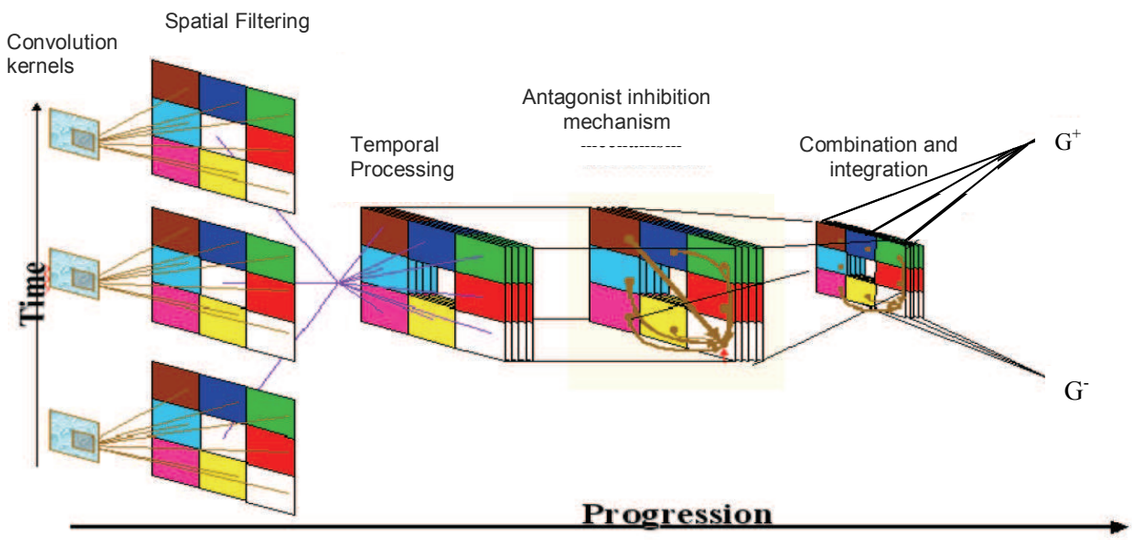


Figure 3. CONEPVIM model. It is divided in four stages: the spatial treatment of the images, the temporal processing of the information (these both stages are grouped in CSTF), the antagonist interaction mechanism (AIM) and the combination and integration of the information

Let *I(x, y, t)* be an image sequence representing the shape of intensity in the time-varying image, assuming that every point has an invariant brightness.
Let us assume that $I(x, y, t) = I(x - ut, y - vt)$ where *(u, v)* is the motion vector of a small region f the image, and where *I(x, y)* is the frame of the image sampling the time $t = 0$.

Thanks to the hypothesis of a high enough sampling frequency to ensure local motion detection, we may assume an immediate constant local speed. Therefore, for a given supposed motion direction and speed, we expect to identify a local motion by finding a spatial contrast at expected places and times.
By applying the oriented Gabor filter, $G_\theta(x,y)$ with $0 \leq \theta \geq \pi$, in $I(x, y)$ we obtain (intensity consevation principle):

$$\begin{aligned} D_\Theta(t) &= \iint_{t=0} \frac{dI(x,y,t)}{dt} * G_\Theta(\hat{x}-\hat{u}, \hat{y}-\hat{v})dxdy \\ &= d\iint_{t=0} \frac{I(x,y,t) * G_\Theta(\hat{x}-\hat{u}, \hat{y}-\hat{v})dxdy}{dt} \end{aligned} \tag{1}$$

Where the rotational equations are given by:

$$\begin{aligned} \hat{x} &= (x-\xi)\cos\Theta - (y-\eta)\sin\Theta \\ \hat{y} &= (x-\xi)\sin\Theta - (y-\eta)\cos\Theta \end{aligned} \tag{2}$$

where $(\xi,\eta) \in \Upsilon$ is a small neighbourhood around (x,y) and:

$$\begin{aligned} \hat{u} &= \frac{t-t'}{\mathrm{T}-1} v\cos\Theta \\ \hat{v} &= \frac{t-t'}{\mathrm{T}-1} v\sin\Theta \end{aligned} \tag{3}$$

for T consecutive images, $t' \leq t$, and the supposed velocity v that ranges from $-w$ to w, where w is the number of supposed absolute speeds.
$(\hat{x},\hat{y})$ is the place where the oriented Gabor signal is going to be computed in a standard way:

$$G_\Theta(\hat{x},\hat{y}) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left(-\frac{\hat{x}^2}{2\sigma_x^2} - \frac{\gamma\hat{y}^2}{2\sigma_y^2}\right)\exp(2\pi_i\frac{\hat{x}}{\lambda} + \phi) \tag{4}$$

Is the response function to the impulse of the Gabor filter that models the function of the ganglion magnocellular cells, where γ is the eccentricity of the receptive field and σ_x, σ_y the dimensions, λ is the wavelength and ϕ is the phase Y. Discretizing the equation 4, we finally compute the following spatial-temporal filter:

$$f_{\mathrm{T},\Theta,v}(x,y,t) = \sum_{t'=0}\sum_{\hat{x},\hat{y}} G_\Theta\left(\hat{x} - \frac{t-t'}{\mathrm{T}-1}v\cos\Theta, \hat{y} - \frac{t-t'}{\mathrm{T}-1}v\sin\Theta\right) \tag{5}$$

However, the measure obtained by a single filter is not ale to determine the 2D motion vector. It is necessary to use a set of filters that differ only in moition. Then they are gathered in a vector called motion sensor vector where every orientation is a motion sensor.

3.2 Antagonist interactions mechanism (AIM)

The second stage of model describe in [Castellanos-Sánchez et al., 2004] (depticted in the centre of figure 3) emulates an antagonist interaction mechanism by means of excitatory-inhibitory local interactions in the different oriented cortical columns of V1.

In this mechanism each neuron receives both excitation and inhibition signals from neurons in a neighbourhood or influence range to regulate its activity. The figure 3 shows the excitatory and inhibitory local interactions where neurons interact with their close neighbours in this mechanism that change the internal state of neurons and, consequently, their influence range, which generate a dynamic adaptive process. The exitatory and inhibitory influence ratios are defined as $\Omega^{\Theta_E}_{(x,y)} = \{(\xi_e, \eta_e) | \ | x - \xi_e | \leq \xi_E, | y - \eta_e | \leq \eta_E, \xi_e, \eta_e \in Z\}$, where ξ_E and η_E are the superior limits of the exitatory influence ratio. And the inhibitory influece ration given by $\Omega^{\Theta_I}_{(x,y)} = \{(\xi_i, \eta_i) | \ | x - \xi_i | \leq \xi_I, | y - \eta_i | \leq \eta_I, \xi_i, \eta_i \in Z\}$, where ξ_I and η_I are the superior limits of the inhibitory influence ratio. This is done for each one of the orientations.

Usually in excitatory-inhibitory neural models, the weighted connections to and from neurons have modulated strength according to the distance from one another. Nevertheless, we call it an interaction mechanism because the inhibitory connections among neurons regulate downwards the activity of opposing or antagonist neurons, i.e. neurons that do not share a common or similar orientation and speed. On the other hand, excitatory connections increase the neuron activity towards the emergence of coherent responses, i.e. grouping neuron responses to similar orientations and speeds through an interactive process.

Then the updaiting of the of the internal state of a neuron is given by:

$$\begin{aligned} \eta \frac{\partial H(x,y,T)}{\partial T} = & -A \cdot H(x,y,T) \\ & + (B - H(x,y,T)) \cdot Exc(x,y,T) \\ & - (C + H(x,y,T)) \cdot Inh(x,y,T) \end{aligned} \tag{6}$$

Where $-A \cdot H(\cdot)$ is the passive decay, $(B - H(\cdot)) \cdot Exc(\cdot)$ the feedback excitation and, $(C + H(\cdot)) \cdot Inh(\cdot)$ the feedback inhibition. Each feedback term includes a state-dependent nonlinear signal, $(Exc(x,y,T)$ and $Inh(x,y,T))$ and an automatic gain control term ($B - H(\cdot)$ and $C + H(\cdot)$, respectively). $H(x,y,T)$ is the internal state of the neuron localised in (x,y) at time T, $Exc(x,y,T)$ is the activity due to the contribution of excitatory interaction in the neighbourhood $\Omega^{\Omega_E}_{(x,y)}$ and $Inh(x,y,T)$ is the activity due to the contribution of inhibitory interactions in the neighbourhood $\Omega^{\Omega_I}_{(x,y)}$. Both neighbourhoods depend on the activity level of the chosen neuron in each direction. A, B and C are the real constant values and η is the learning rate. For more details on the excitation and inhibition areas see [Castellanos-Sánchez et al., 2004; Castellanos-Sánchez, 2005].

The excitation is defined as follows:

$$Exc(x,y,T) = H(x,y,T) + \sum\nolimits_{\Omega^{\Theta_E}_{(x,y)}} W_{\Theta_E}(x,y) L(x,y) \tag{7}$$

and the inhibition is calculated by

$$Inh(x,y,T) = \sum_{\Omega^{\Theta_I}_{(x,y)}} W_{\Theta_I}(x,y) L(x,y) \tag{8}$$

and where

$$W_{\Theta_{E,I}}(x,y) = \begin{cases} g(d(x,y,\xi_e,\eta_e), R(x,y,t), \mu, \sigma) & \text{if } d(\cdot) < R(\cdot) \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

with $g(d(\cdot), R(\cdot), \mu, \sigma)$ as a gaussian centered in (x, y) with mean μ and standar deviation σ. Let $R(x,y,t)$ be the influence ratio of neuron (x,y) defined as $\Gamma | H(x,y,t)/saturation$, where Γ is the proposed influence ration and $saturation = 2\max_{(x,y,t,\theta,v)}(H(x,y,t))$. This neuron receives at most $R(x,y,t)^2$ excitatory connections from neurons with the same direction and speed and at most $(V \cdot \Theta - 1) \cdot R(x,y,t)^2$ inhibitory connections from other close neurons. At this level, each pixel correspond to $\Theta \cdot V$ different neurons that encode information of directions and speeds. Finally $L(x,y)$ is the algebraic sum on the all speeds, for each orientation.
The computations described in this subsection analysing its neural and synaptic parallelism have been implemented on FPGA circuits [Girau et al., 2005].

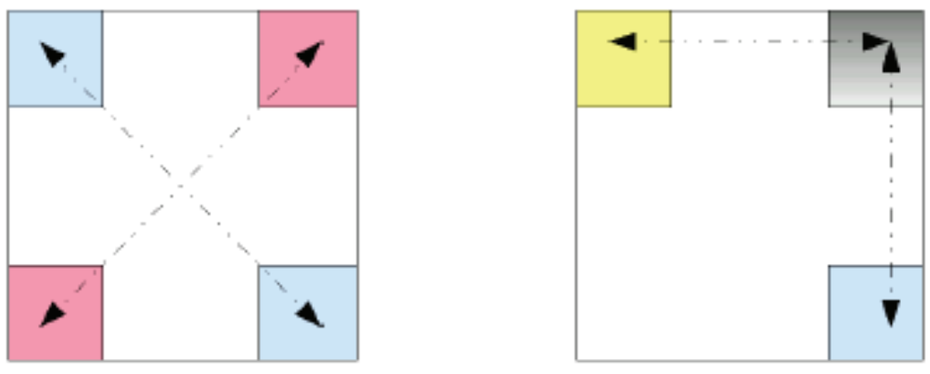


Figure 4. Different directions of controlled sub sequences of real images generated for each supposed speed

4. Neuromimetic indicators

The visual perception of motion is not totally determined in the local responses of the V1 neurons. They are processed to obtain the speed after being collected and combined from V1 and being integrated in MT. It is this combination of signals that resolve the local ambiguity of responses of neurons in V1 [Castellanos-Sánchez, 2005]. This activity is the inspiration of the last part of figure 3 (directions and speeds combination and integration).

4.1 Controlled generation of sequences of real images

The model described here has several parameters to be fixed. The results shown are the product of several experiments. To begin with, we analysed the active neurons in each direction and speed, the frequencies of active neurons after updating (ANaU) and the negative updating increase (NUI) through m different sequences of real images about 384x288 pixels per image.

Next, to analyse ego-motion, we selected n images of each sequence of real images and for each selected image we generated $\Theta \times V$ controlled sub-sequences (Θ are different directions and V different speeds) indicated in the figure 4.

Finally for motion classification, we took a subsequence of each sequence of real images too where : a) the motion does not exist, b) one object moves, c) two or more objects move simultaneously. The interpretation of the different obtained values are shown in the next subsections.

4.2 Motion type

It is important to mention that this neuromimetic indicator is purely based on the interactions between cells of the orientation columns in V1 and integration and treatment of the proposed velocities in the MST cells.

The equation 6 shows the actualisation rule in the AIM for the active neurons. Let S be a real image sequence and let $R \subset S$ be a subsequence with $Card(R) = \tau$ the subsequence size and let p be the percentage of the neurons to update.

The AIM mechanism updates $p\%$ of active neurons and we obtain in it two frequency percentages : the active neurons after updating ($ANaU$) and negative updating increase (NUI, see the right side in the equation 6).

The frequencies of the products between $ANaU$ and NUI indicators in all the different controlled sub-sequences (see last section) inspire us to propose our first neuromimetic indicator: **neuromimetic motion indicator**, $NMI = ANaU * NUI$. The experimented ranges of NMI obtained are shown in table 1.

4.3 Speed and direction

Once the orientations have been estimated by the V1 cells the information must be collected, integrated and homogenized in MT, following a hierarchical principle (grouped by orientation columns). Achieving this way a disambiguation of the orientation (caused by the local treatment in V1 cells) of the moving objects, and so forth, it solves the local aperture problem and homogenates the orientations. In the case of the speed the neurons in MST group the information from the different speeds, using a extended-type of receptive fields, to offer an estimation of the velocity.

The equation 6 shows the actualisation rule in the AIM for the active neurons. Let S be a MT neurons sum the responses of V1 neurons with receptive field positions inside a local spatial neighbourhood that is defined through time and generates a response according to the speed of the visual stimulus [Castellanos-Sánchez, 2005]. This locality of the AIM mechanism on all the several considered motion directions in V1 bring an emerging answer corresponding to the global direction [Castellanos-Sánchez et al., 2004; Castellanos-Sánchez, 2005].

Condition	Description
NMI < 0.10	Null motion
NMI < 1.00	Small movin objects or bruit
NMI < 5.00	One or two moving objects
NMI < 10.00	Three to five moving objects
NMI < 40.00	Six or more moving objects, or ego-motion
NMI < 250.00	Ego-motion or big moving objects
NMI < 400.00	Ego-motion
NMI ≥ 400.00	Strong Ego-motion

Table 1. Experimental ranges for neuromimetic motion indicator (NMI)

On the other hand, neurophysiological studies roughly indicate that neurons in MT of the visual cortex of primate brains are selective to speed of visual stimuli; which implies that neurons respond strongly in a preferred direction and with a preferred speed [Simoncelli, 1998].

For each real subsequence R and for the filtering images generated in the equation 1 we define

$$sat^{+} = \max_{t,\Theta,v}(H_{t,\Theta,v}(x,y)), sat^{-} = \min_{t,\Theta,v}(H_{t,\Theta,v}(x,y)) \tag{10}$$

where sat^{+} and sat^{-} are the positive and negative saturation respectively.

For each direction and speed of each neuron, we count the neurons with a response greater than at. This parameter is the average of positive and negative saturations. The equation 12 shows its behaviour and the equation 11 computes this frequency in direction θ with speed v.

$$C(\Theta,v) = \sum_{x,y} D(at, H_{t,\Theta,v}(x,y)) \tag{11}$$

with

$$D(at, H_{t,\Theta,v}(x,y)) = \begin{cases} 1 \text{ if } H_{t,\Theta,v}(x,y) > at \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

where $D(\cdot,\cdot)$ is the threshold of the CSTF filtering.The collection and combination in MT for direction estimation is computed by:

$$E(\Theta,v) = 3 \cdot C(\Theta,v) + 2 \cdot (C(\Theta-\phi,v)) + C(\Theta+\phi,v)) + C(\Theta-2\phi,v) + C(\Theta+2\phi,v) \tag{13}$$

where $\phi = 2\pi/\Theta$ is the separation in degrees between each oriented column and $E(\cdot,\cdot)$ is the sum of several oriented responses of V1 that activate a neuron in MT. Finally we computed the frequencies for negative and positive supposed speeds by the following equations:

$$G^{+} = \sum_{v>0,\Theta} C(\Theta,v), \qquad G^{-} = \sum_{v<0,\Theta} C(\Theta,v) \tag{14}$$

Then we arranged $E(\Theta,v)$ in a direction according to each speed and arranged G^{+} and G^{-} too for processing them to obtain speed and direction indicators. These indicators will be describe in the next two paragraphs.

4.3.1 Speed

To obtain the winner speed, we propose the **neuromimetic speed indicator (NSI)** defined by following equation:

$$NSI = \frac{100 \cdot \min(G^+, G^-)}{\max(G^+, G^-)} \tag{15}$$

With this indicator we compute the relative speed (*rs*) that compares the different speed frequencies and their proportion. The table 2 shows our experimental values for $V=5$. Then $v\_i=\{-2,-1,0,1,2\}$, with $v1$ is the frequency of $|v\_i|=1$ and $v2$ is the frequency of $|v\_i|=2$.

4.3.2 Direction

Finally, for an interpretation of integration of directions for each neuron in MT, we compute the equation 10 for each direction and speed. Next, we arrange their values from major to minor and we take the first three. If these candidates are contiguous in direction, the winner will be at the centre of the three candidates' directions. This is our **neuromimetic direction indicator *NDI***.

Finally, if the maximum of the two computed speeds in the equation is the negative one, the winner direction will be its antagonist, ei, $\Theta = \Theta - 180^o$.

Type	Condition Relative speed	Prototype speed
Weak if $v1 > v2$	NSI > 70.0 rs = (100.00-NSI)/29.0	0
	NSI > 12.0 rs = (71-NSI)/59 + 1	1
	otherwise rs = (12-NSI) * 0.3529 / 12 + 1	2
Strong if $v1 < v2$	NSI > 22.0 rs = (NSI * 0.6470)/22 + 2.3530	3
	NSI > 39.0 rs = (NSI - 22) / 10 + 3	4
	Otherwise Speed not processed	≥ 5

Table 2. Experimental ranges for neuromimetic speed indicator (NSI)

5. Results

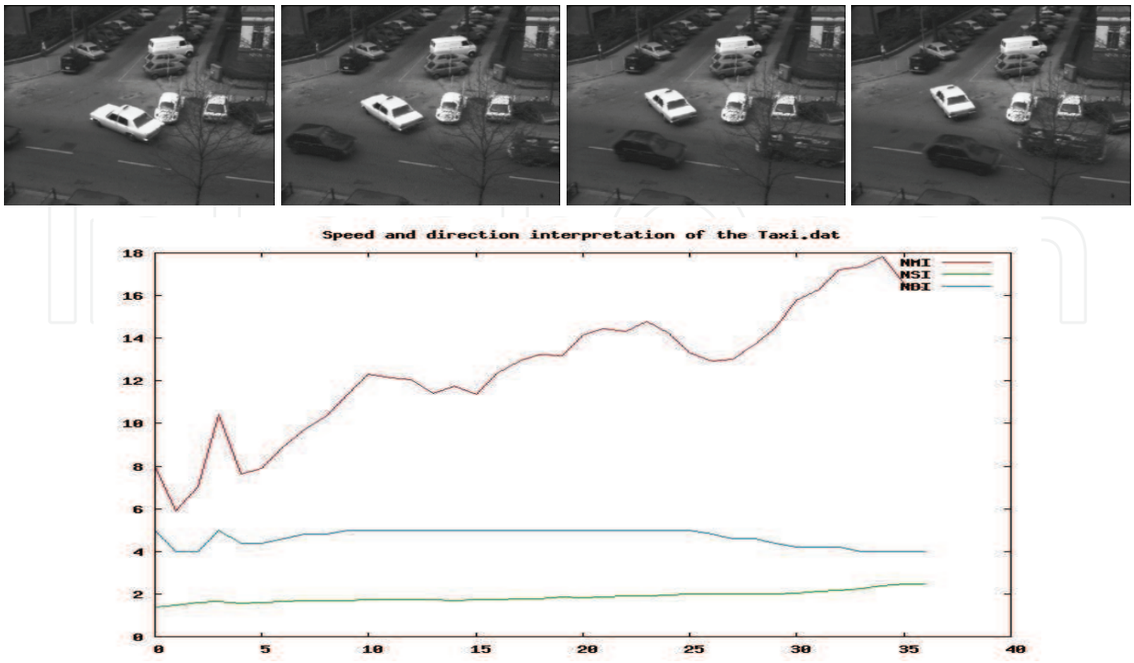
The free parameters of our model were set according to the suggestions in [Castellanos, 2004]. We chose only three sequences of images among m = 50 analysed sequences : the Yosemite Fly-Through (sequence of synthetic images), the Hamburg Taxi, the Karl-Wilhelm (DKW, traffic video surveillance) and the BrowseB (issue of a surveillance camera). They include various numbers of RGB images (15, 42, 1035 and 875 images, respectively) and of sizes of : 316×252, 256×191, 702×566, 384×288, respectively, and they are first gray-scaled.

The figure 5 and 6 shows four images of these sequences and their graph of the proposed neuromimetic indicators. The values of NMI are between 0 (null motion) and 1000 (ego-motion), of NSI between 0 and 6, and NDI is in {1, 2, 3, 4, 5, 6, 7, 8} (0°,45°,...,315°).

The real Hamburg Taxi sequence shows three moving cars and a pedestrian. The NMI is between 6 and 18, then according to the table 1 there are about three moving objects and the global speed is 2 pixels per image moving at approximately 180° and end at around 135°.

The BrowseB sequence issue of video surveillance in the hall of INRIA laboratory, Grenoble, France, may be split into three parts : (1) a person walks to the centre, stops and returns; (2) there is no motion; (3) another person walks in, stops and goes farther.

The Hamburg Taxi Sequence.



The BrowseB sequence.

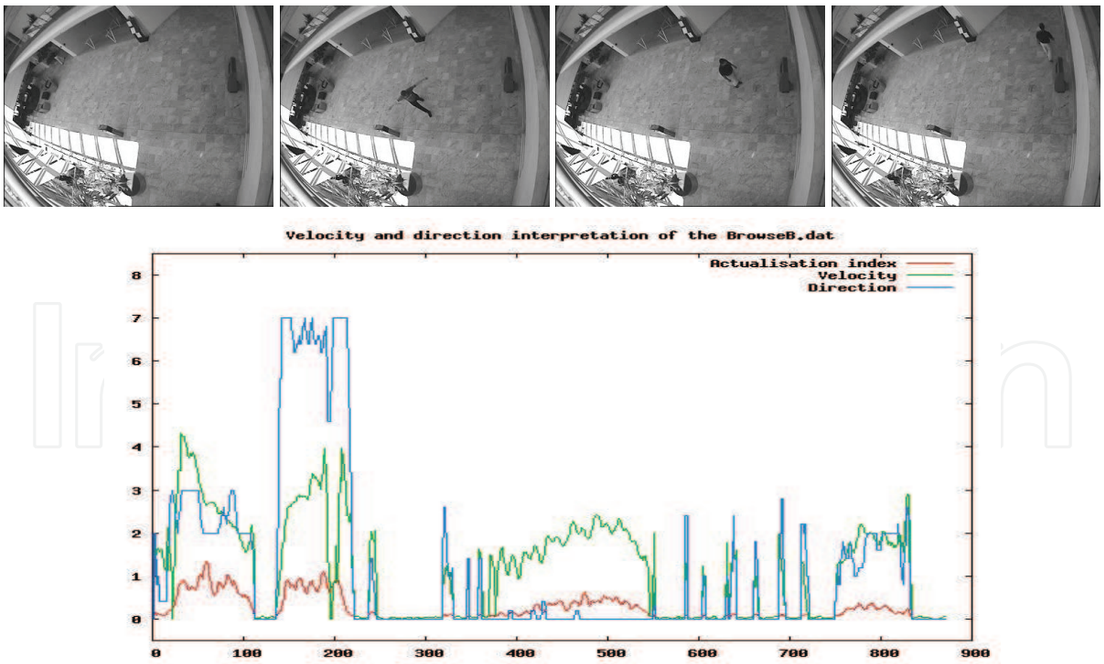
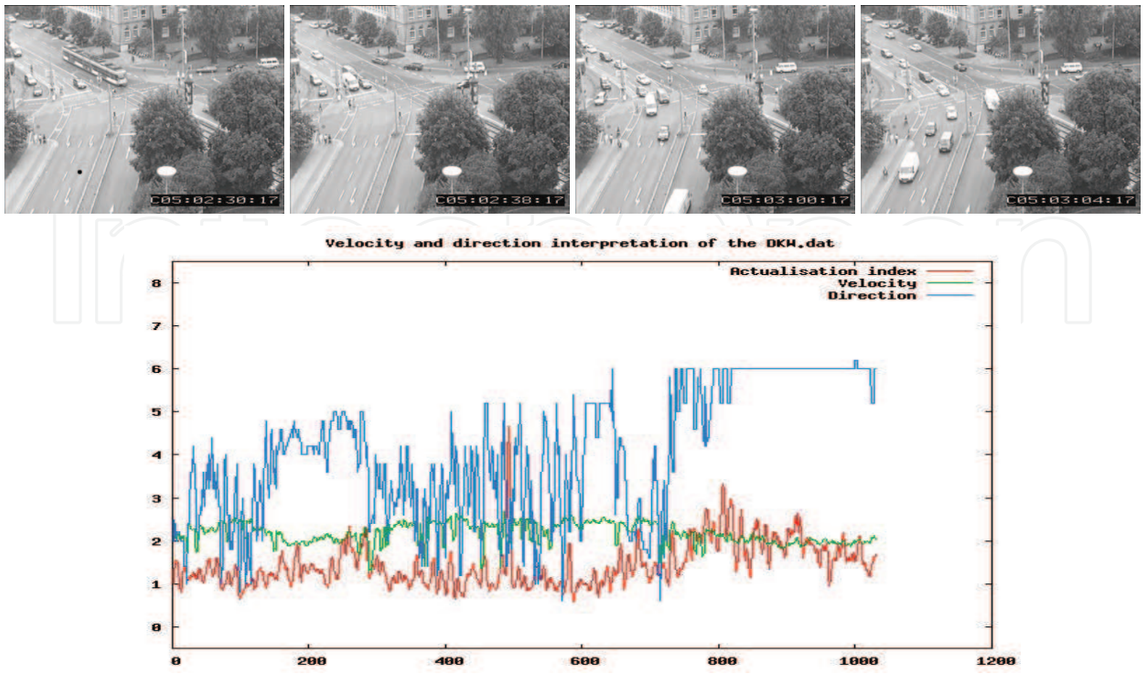


Figure 5. Two natural sequences, showing the sequence contents. In the first line and down of each one the graphic describing its motion behaviour (decomposed in the three neuromimetics indicators)

The DKW sequence



The Yosemite Sequence

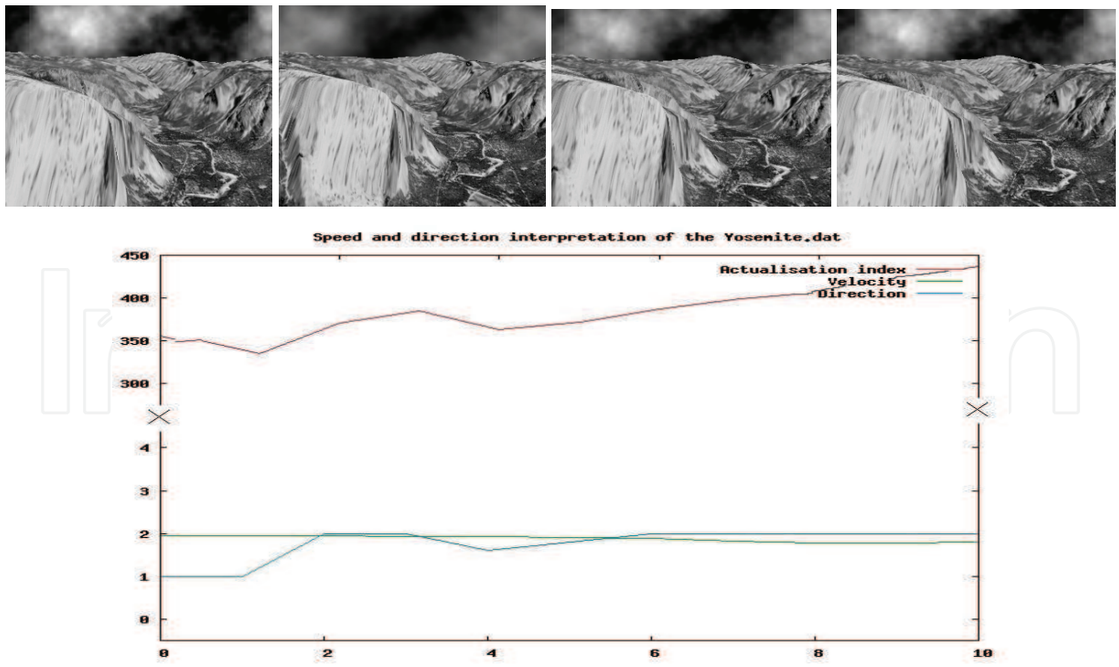


Figure 6. Sequences of real and synthetic images, Yosemite and DKW sequence below each one the graphic describing its motion behaviour (decomposed in the three neuromimetics indicators)

For the case of the BrowseB sequence in figure 5. The first part (images 0 to 220) may be split into three parts according to NMI : two parts with motion and the other part with null motion that correspond to the first person walking between 90° and 135° and with a speed of 4 to 2 pixels per image, stops and returns between 270° and 315° and with a speed of 2 to 4 pixels per image. For the second part (images 221 to 325) there is null motion. The last part may be split too into three parts according to NMI : (1) motion, (2) generally null motion and (3) motion, respectively to describe this part of the BrowseB sequence. The person walks approximately at 0° with a speed of 1-2 pixels per image. Next, a period of null motion with very weak motions (see pics in the graph between image 550 and 750). Finally, the person moves to about 90° with a speed of about 2 pixels per image.
The DKW sequence shows the images taken with a surveillance camera. In the first part (images around the 230) describe an increment in the motion due to the amount of objects moving in the scene. This kind of increments are accented in images around the 700, where after a period of low motion (images from 580 to 600). Besides the speed tends to remain stable during the whole sequence, this because the combination of objects in motion is almost the same for the whole sequence. Besides, in this last sequence it is important to mention that the obtained results of the direction are due to the combination of the objects that are moving in multiple direction (south or north, east or west).
Finally, the synthetic Yosemite Fly-Through sequence shows an aeroplane flying on the mountains, and mainly presents an ego-motion with a speed of five pixels (down image) that diverge and two pixels for the moving clouds to the right (top image). The NMI is between 300 and 450, then according to the table 1 it proposes an ego-motion with 2 pixels per image moving at around 45°.
In the figure 7, we show the average of direction and speed of optical real flow of Yosemite sequence and the experimental results obtained by our model. Our model presents a conceptual error about 22.5°, despite which it is sufficient to describe the real movement towards the North-East. Finally, the speed is not numerically exact, but our estimation is very similar to the real one. Then, the global motion obtained here is very similar to the Yosemite Fly-Through data.

6. Conclusions and Perspectives

This work is based on the CONEPVIM model [Castellanos-Sánchez et al., 2004]: a neuromimetic connectionist model for visual perception of motion. A model fully inspired by the visual cortex system, the superior areas and their relations.
In this paper we took advantage of the low-level analysis to detect local motions to obtain the global speed and direction. They are determined by the neuromimetic motion indicator issued by AIM mechanism.
Our first experiments show that this model is capable of estimating the null motion, simple motion and ego-motion with an estimation of global speed and direction in an environment where other persons or objects move. The estimation of motion is robust in quite complex scenes without any predefined information. Nevertheless, the estimation of NMI is fastidious. The experimental values are correct for the sequence of real images of ±33% the size of 384 × 288.
Besides, it has been seen that the proposed model gives good results, it has do basically with two characteristics, the first one is the local and distributed treatment of the information used for the analyse of the sequences. The second is the integration of these stages of processing, the last one is very important due to it is strongly linked to the methodology proposed to overcome the task.

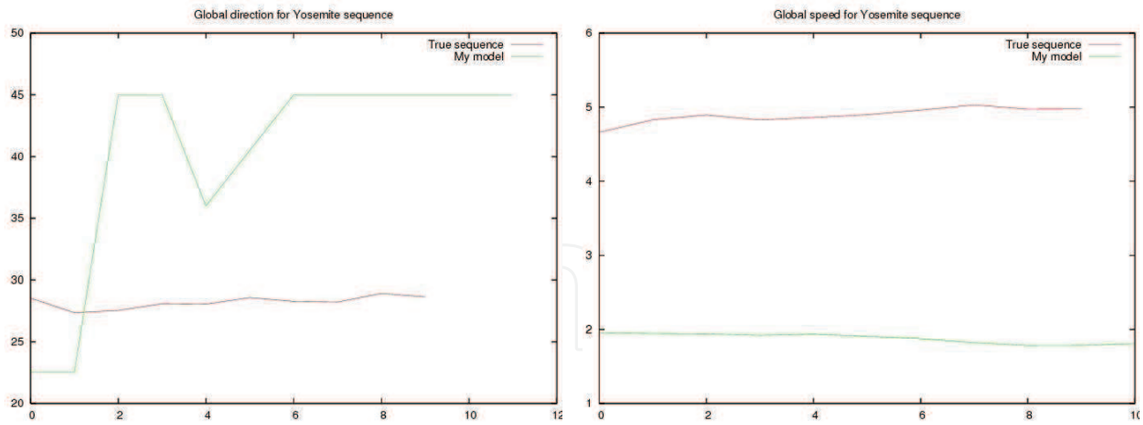


Figure 7. Comparison between the optical real flow of Yosemite sequence and the experimental results obtained by our model. In this case, both the speed and direction estimated are very similar to the results that are taken as the true. In the case of the direction the precision we are handling (the separation of the groups of neurons in 45^o sets) gives a range of error

In this sense we have some perspectives about the methodology:

1. Following this methodology it is possible to understand not only the type of motion that is perceived but also to understand what is moving in the scene, since this processing belongs to the Ventral Pathway and in theory it works very similar to the Dorsal Pathway. This because it might be inferred that the processing in this path might also have strong interactions with the Dorsal Pathway, which help us to depict the information perceived.
2. Assuming that the information from other senses (auditory or sensitive) in the brain is processed by groups of neurons (just like it happens in the case of the visual processing) so this methodology will be helpful to understand how these arrangements of neurons work and how the information is combined with the information from other senses to interact with the environment.

8. Acknowledgment

This research was partially funded by project number 51623 from "Fondo Mixto Conacyt-Gobierno del Estado de Tamaulipas".

8. References

Adelson E. H., Bergen J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2(7):284–299.

Castellanos-Sánchez C. (2005), Neuromimetic connectionist model for embedded visual perception of motion. *PhD hesis*, Université Henri Poincaré (Nancy I), Nancy, France, Biblioth`eque des Sciences et Techniques.

Castellanos-Sánchez C., Girau B., Alexandre F. (2004). A connectionist approach for visual perception of motion. In Smith, L., Hussain, A., Aleksander, I., eds.: *Brain Inspired Cognitive Systems* (BICS 2004). BIS3–1:1–7.

Derrington A. M.,Henning. G. B.(1993). Detecting and discriminating the direction of motion of luminance and colour grattings. *Visual Research*, 33:799–811.

Fellez, W.A., Taylor, J.G. (2002). Establishing retinotopy by lateral-inhibition type homogeneous neural fields. *Neurocomputing* 48. 313–322.

Huitzil-Torres C., Girau B., Castellanos-Sánchez C. (2005). On-chip visual perception of motion: A bio-inspired connectionist model on fpga. Pages 557-565.

Grzywacz N. M., Yuille A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. In *Proceedings Royal society London* B, volume 239, pages 129–161.

Grzywacz N. M., Yuille A. L. (1991). Theories for the visual perception of local velocity and coherent motion. In *Computational models of visual processing*, pages 231–252. M. I. T.

Heguer D. (1987). Model for the extraction of the image flow. *Journal of the Optical society of America*. 1455-1471.

Hengyi R., Tiangang Z., Yan Z., Silu F., and Lin C.(2003). Spatiotemporal Activation of the Two Visual Pathways in Form Discrimination and Spatial Location: A Brain Mapping Study. *Human Brain Mapping*, 18:79–89.

Hubel D. H., Weisel T. N.. (1962). Receptive fields, binocular interaction and functional architecture in the cats visual cortex. Journal Physiology, 160:106–154,.

Imbert M. (1983). Neurobiology of the Image. *La Recherche*, 144:600-613.

Jones J. and Palmer L. (1987). An evaluation on the two dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1233–1258.

Marcelja S. (1980). Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America* A, 70/11:1297–1300.

McCane, B., Novins, K., Grannitch, D., Galvin, B. (2001). On benchmarking optical flow. *Computer Vision and Image Undestanding*.126–143.

Mingolla, E. (2003). Neural models of motion integration and segmentation. *Neural Networks* 16. 939–945.

Moga, S. (2000). Apprendre par imitation: une nouvelle voie d'apprentissage pour les robots autonomes. *PhD thesis*, Université de Cergy-Pontoise, Cergy-Pontoise,France.

Pack, C., Grossberg, S., Mingolla, E. (2000). A neural model of smooth pursuit control and motion perception by cortical area MST. *Technical Report* CAS/CNR-TR-99-023, Department of Cognitive and Neural Systems and Center for Adaptive Systems, 677 Beacon St, Boston, MA 02215.

Pollen D., Ronner S. (1981). Phase relationships between adjacent simple cells in the visual cortex. *Science*, 212:1409– 1411, 1981.

Sekuler R., Watamaniuk S. N. J., Blake R. (2002). Motion Perception. *Steven's Handbook of Experimental Psychology*, 1:121–176.

Simoncelli, E.P., Heeger, D.J. (1998). A model of neural responses in visual area MT. *Vision Research* 38. 743–761.

Van Santen J. P. H., Sperling G. (1984). Temporal covariance model of human motion perception. *Journal of the Optical Society of America* A, 1:451, 1984.

Van Santen J. P. H., Sperling G. (1985). Elaborated Richardt detectors. *Journal of the Optical Society of America* A, 2:300– 321, 1985.

Zemel, R.S., Sejnowski, T.J. (1998). A model for encoding multiple object motions and self-motion in area MST of primate visual cortex. *The Journal of Neurosciences* 18. 531–547.

Brain, Vision and AI

Edited by Cesare Rossi

ISBN 978-953-7619-04-6

Hard cover, 284 pages

Publisher InTech

Published online 01, August, 2008

Published in print edition August, 2008

The aim of this book is to provide new ideas, original results and practical experiences regarding service robotics. This book provides only a small example of this research activity, but it covers a great deal of what has been done in the field recently. Furthermore, it works as a valuable resource for researchers interested in this field.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Claudio Castellanos Sanchez and Pedro Luis Sanchez Orellana (2008). Bio-inspired Connectionist Modelling: An Application to Visual Perception of Motion, Brain, Vision and AI, Cesare Rossi (Ed.), ISBN: 978-953-7619-04-6, InTech, Available from: http://www.intechopen.com/books/brain_vision_and_ai/bio-inspired_connectionist_modelling__an_application_to_visual_perception_of_motion

INTECH

open science | open minds

InTech Europe

University Campus STeP Ri

Slavka Krautzeka 83/A

51000 Rijeka, Croatia

Phone: +385 (51) 770 447

Fax: +385 (51) 686 166

www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai

No.65, Yan An Road (West), Shanghai, 200040, China

中国上海市延安西路65号上海国际贵都大饭店办公楼405单元

Phone: +86-21-62489820

Fax: +86-21-62489821

IntechOpen