NETWORK MECHANISMS FOR RETINAL MOTION DESCRIPTION SUGGESTED BY A BIO-INSPIRED ARTIFICIAL VISION SYSTEM

Roberto Moreno-Díaz jr, Alexis Quesada-Arencibia,
José Carlos Rodríguez-Rodríguez and Gabriel de Blasio

Received January 18, 2006

Abstract. We present the results of a computer vision system intended to describe in real time the trajectories of moving objects in a variety of speeds. The developed computing mechanism includes channel processing for instant position and velocity estimation, while trajectory plotting is made by combining direction and speed of movement. Retinal ganglion cell-like computational elements are randomly distributed thorough the input image, being direction-selective, allowing lateral interactions among them to correct peep-hole effects in movement analysis, presenting a variable degree of overlapping of their receptive fields (RF). We show results and errors of the working system when all parameters of cells are varied. One unexpected characteristic is that of the degradation of results in trajectory estimation when the size of receptive fields or the number of receptive fields, i.e. of computing cells, increase. In order to get good results for a variety of moving object speeds we have to reach a compromise between size and number of cells. This effect has interesting implications when translated back to the biological system which originally inspired the design of the computational method and could be used as a rationale to explain the distribution, morphological characteristics and number of movement detecting neurons.

1 General system description, implementation and results The study of biological basis of motion detection and analysis and the modelling and artificial implementation of those mechanisms has been a fruitful path of science in the last 60 years. When trying to implement an artificial vision system that has to work in very demanding conditions (e.g. real time, rapid changing visual environment etc) we found some of the biological concepts quite useful, thus is the case with receptive fields, overlapping, lateral interactions of computing units etc. [5], [6], [7]. However, some traditional approaches of visual computing (formal mathematical analysis, strict geometric patterns of neuron-like processors, selectivity of stimulae) have represented an extra restriction affecting to the overall performance of the system. After questioning these traditional approaches and including some "messy" characteristics of real biological systems such as random distribution of neuron-like processors, non homogeneity of neural architecture, channel processing, sudden failure of processing units and, in general, non deterministic behaviour of the system we found and show how reliability of motion analysis, computational cost and extraction of pure geometrical visual descriptors (size and position of moving objects) are improved in an implemented model.

1.1 Image acquisition and control of moving objects The system has been implemented on a Pentium$^{\mathrm{TM}}$PC type computer in which an acquisition and digitation board (IC-ASYNC, Imaging Technology$^{\mathrm{TM}}$) has been installed. This image board processes b/w

2000 *Mathematics Subject Classification.* 68U10, 92-08, 92B20.
Key words and phrases. Neuron-like computation, channel processing, performance optimization, visual motion description.

images, so a monochromatic progressive camera CV-M10 with 640x480 pixels resolution is used, though the recorded images are 256x256 pixels and 256 grey levels. Software has been developed under Visual Basic 5.0. All artificial vision systems being extremely sensitive to illumination conditions, our image acquisitions take place in the close, strictly controlled structure. This structure has a working surface of aprox. 1 m2, and is 2m high. The camera is on top of the structure, located at 1.5 m from the surface where the object is moving. This object, square shaped, is built using Lego's™Robotics Invention System©, a construction kit that includes a programmable microprocessor and several sensors (light, touch, rotation) and motors that allows us to program, control and monitor its behaviour.

To detect motion, a sequence of images of the moving object that follow one after the other is acquired and processed in its temporal order. However, the same kind of series of receptive fields that have been generated may also be used to estimate the movement of an object into the retina. To detect that an object is moving, the receptive fields working as contrast detectors locate the borders of the object and then follow their movement. In case of having a textured object moving across the retina, not only the borders of the object, but also its contrast areas would be useful to identify its motion, e.g., the zones where the grey level changes are followed in time and relocated to obtain an estimation of motion.

1.2 Direction selectivity and speed estimation. The randomly generated receptive fields are now divided into 8 classes, one for each of the equidistant directions considered, with an equal number of receptive fields for each one as we suppose that none of them is more important for the system than the others.

The origins of the receptive fields are calculated only once for the eight directions, so that each one has eight receptive fields associated as in a star, that act independently to locate motion in one of the eight directions described above. When the origin of one of the receptive fields detects a border, the weights are changed in the following way to find out the direction and speed of the movement (Figure 1).

The first values allow us to locate the boundaries or changes when they pass over the origin of the receptive fields, while the second ones make it possible to relocate the boundary or change if it moves in the preferred direction of the receptive field, once it has arrived to the origin. Although the eight receptive fields are represented here as a whole, they act independently, once the border or contrast area of the object is located, so that, as the object moves, some of them (none, one or more than one) react, depending on the direction of its motion. With these new weights, if the border moves in the preferred direction, the receptive field will react. These are the direction selective cells of the system.

In order to be able to discriminate not only the direction but also the speed, the response of the receptive fields is divided into two components, one related to direction itself, that will be 1 or 0, depending on the excitation of the receptive field by a border or change moving in its preferred direction, and another one related to speed, that will be 1,2,3,... according to the distance of the border from its position in the previous instant.

Thus, the larger the receptive fields are, the faster the maximal measurable speed will be, since the distance into the receptive field that the object can cover in one step is longer (in the example of figure 1, this distance is 3 pixels/ time step). As the outputs of the receptive fields are divided into two components, each direction has two summing boxes, one for direction itself, and another one for speed that receive the outputs of all the receptive fields specialized in that direction. The direction that the object follows can be determined by that whose direction summing box has the highest value, since that means that there are more points moving this way as any other one. Similarly, the speed can be obtained by averaging the speed of the points moving in this direction, that is to say, by dividing the value of the speed summing box by the value of the direction summing box for the direction

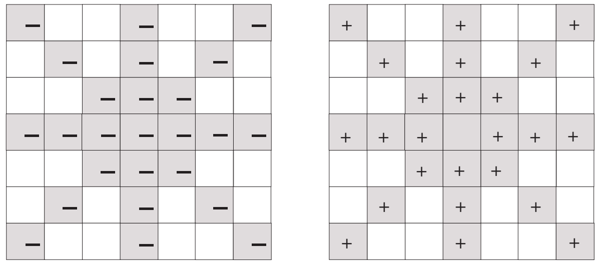

Figure 1: Change of the weights

previously selected (Figure 2). The different shapes that the object may have and the simplicity of the receptive fields causes that a movement in a certain direction also activates many receptive fields specialized in the 'neighbour' directions, thus making it possible to obtain a false or doubtful result in certain instances.

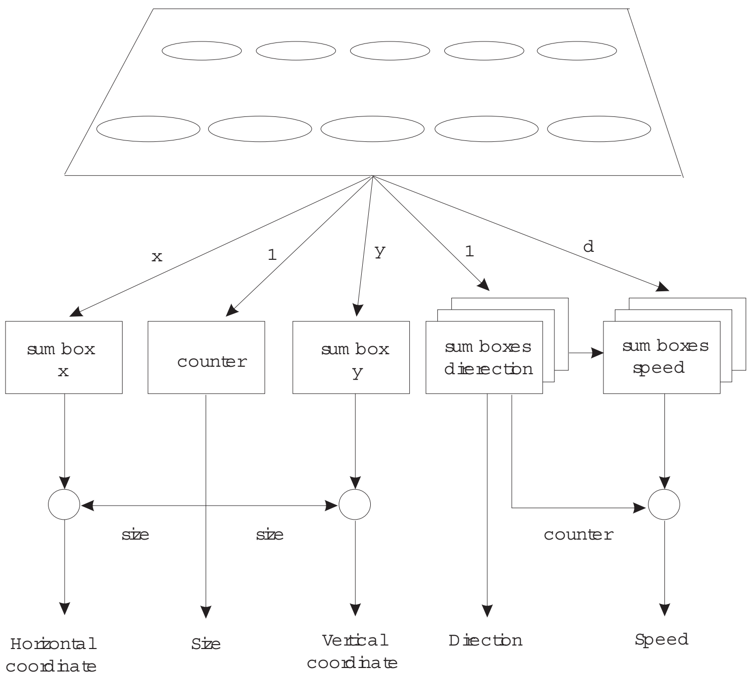

Figure 2: System general diagram

1.3 Movement description and trajectory refining Although that also happens in the natural systems, as cells react not only to a very specific direction, but also to a whole range of directions around the preferred one, it represents a problem when trying to determine approximately the direction of motion with a certain degree of accuracy.

To solve this problem, two different operations are carried out. First, the response for each direction is integrated in time, that is to say, the value associated to each instant is obtained by the sum of the response in the last 3 instants, thus making isolated wrong results disappear. Second, the value obtained for each direction is subtracted from the values obtained for those directions located two places clockwise or two places counterclockwise, in such a way that the neighbours of the real direction undergo a greater decrease than this one.

This time integration makes the instantaneous false results disappear and this kind of lateral inhibition enhances the values of the right direction and decreases those of the neighbour directions thus making it easier to distinguish it among the group of directions activated by the motion of the object. Once the direction of the object has been found, its speed is calculated by dividing the total number of pixels that the borders have covered in that direction by the total number of borders that have moved in the same direction.

2 Results for speed estimation and trajectory plotting The systems was put to work by analyzing video films of the movement of an object following several simple trajectories at a variety of speeds and making an exhaustive study of the influence of the several possible parameters involved (number of computing units, size of receptive fields, effect of lateral inhibition, facilitation and temporal integration). Some of the results for speed estimation related to receptive field size are shown in the following tables (tables 1 and 2), for a vertical trajectory of the object.

Size in Pixels	RFS 3 pixels	RFS 5 pixels	RFS 7 pixels	RFS 10 pixels	RFS 15 pixels	RFS 20 pixels
Estim. Speed (cm/sc)	6,83	6,96	6,96	6,96	6,96	6,96
% Error	-0,29	+1,60	+1,60	+1,60	+1,60	+1,60

Table 1: Speed estimation related to receptive field size. Real object speed: 6,85 cm/s; number of RF: 34992

Size in Pixels	RFS 3 pixels	RFS 5 pixels	RFS 7 pixels	RFS 10 pixels	RFS 15 pixels	RFS 20 pixels
Estim. Speed (cm/sc)	7,56	15,68	19,78	22,57	22,59	22,59
% Error	-66,60	-30,74	-12,63	-0,30	-0,30	-0,30

Table 2: Speed estimation related to receptive field size. Real object speed: 22,64 cm/s; number of RF: 34992

It can be seen that the estimated speed depends on the size of receptive fields: small receptive fields work better for small speeds (in the sense that a stable estimation is reached without greatingly increasing the size) and larger receptive fields for higher velocities. Making the same experiments with an oblique movement of the object, the conclusion remains the same as can be seen in tables 3 and 4:

Size in Pixels	RFS 3 pixels	RFS 5 pixels	RFS 7 pixels	RFS 10 pixels	RFS 15 pixels	RFS 20 pixels
Estim. Speed (cm/sc)	7,78	8,23	8,23	8,23	8,23	8,23
% Error	-4,30	+1,23	+1,23	+1,23	+1,23	+1,23

Table 3: Speed estimation related to receptive field size. Real object speed: 8,13 cm/s; number of RF: 34992

Size in Pixels	RFS 3 pixels	RFS 5 pixels	RFS 7 pixels	RFS 10 pixels	RFS 15 pixels	RFS 20 pixels
Estim. Speed (cm/sc)	7,17	18,06	19,56	19,56	19,56	19,56
% Error	-62,65	+5,93	+1,87	+1,87	+1,87	+1,87

Table 4: Speed estimation related to receptive field size. Real object speed: 19,20 cm/s; number of RF: 34992

As before, the estimate of speed stabilises with smaller RF size (5x5) at the lower speed, vs. the larger RF size (7x7) at the higher speed.

To study the effect of the variation of the number of receptive fields, in determining speed and the estimated trajectory we fix a speed of 22,64 cm/sc getting the following results for a number of cells ranging from 46656 (which corresponds to one RF per input pixel) to 100 (0.2% of the previous figure) and for the usual variation of the receptive field size from 3 to 20 pixels/side (Figure 3):

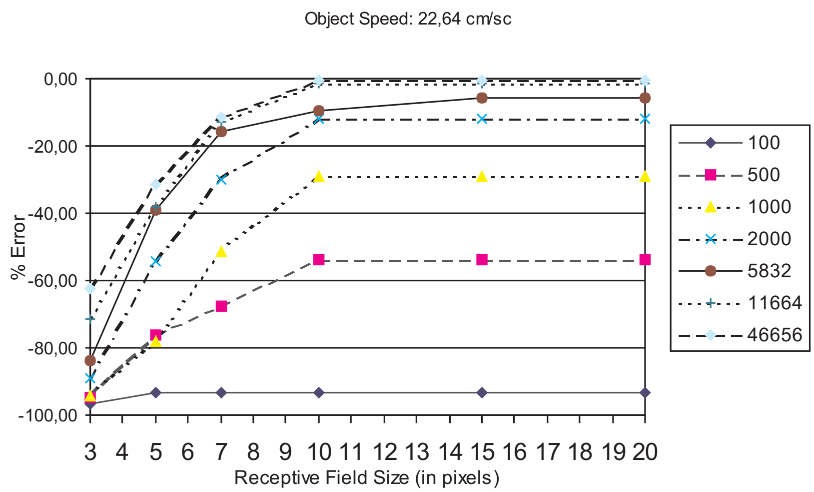

Figure 3: Percent error in the estimated speed vs. Receptive Field Size as function of the number of Receptive Fields.

As can be seen, in order to get a reasonable figure for the estimated speed we need to have at least 2000 receptive fields and a size of at least 10x10 pixels distributed on the retina. Better results are yielded by increasing the number of receptive fields. But with more than 11644 receptive fields there is no practical improvement in the error with a 10x10 pixels receptive field. This suggests that with these numbers of RF the system has a high built-in fault tolerance that allows us to compute values very close to the correct speed estimation (at least with an error margin between 0.31% to 1.37%). At the same time, the computational costs of our system can be controlled. The results in the previous graph refer to a vertical trajectory. The system yields similar results for horizontal and oblique trajectories. With different speeds the number of receptive fields and their sizes for the best estimates will be different.

In the following graphs the plotting of the real trajectory vs the estimated one using only the data from speed and direction channels are shown when using a complex movement.

They graphically show the effect of varying receptive field size with a fix RF distribution of 11664 cells and varying RF field number (keeping them in a random distribution) when a fix size of RF is set to 10 pixels (Figure 4).

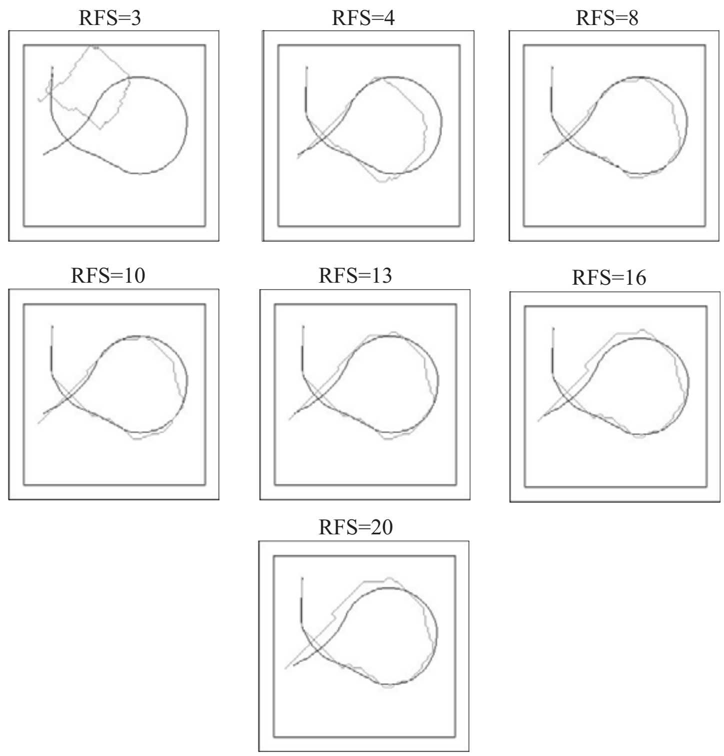

Figure 4: The effect of varying receptive field size with a fix RF distribution of 11664 cells. (RFS=Receptive Field Size expressed in pixels). The solid darker line shows the real trajectory of the object; the lightest shows the calculated trajectory using speed estimation and direction selectivity cells data.

3 Implications in retinal functioning Ganglion cells have been described and studied in all animal retinae, being classified in numerous subtypes that differ in morphology and in their response to visual stimulation. It is widely admitted that they extract different image parameters (or descriptors) and transmit them to the brain in independent parallel channels, [4]. Since the ability to see motion is ubiquitous in visual animals, several of these subtypes are related to movement detection and analysis, and in some animals, like the frog, different movement detection cells are categorized depending also on their responses to more specific characteristics of the moving stimulus, showing that, if necessary, movement is not processed by a unique, differentiated channel but by a number of specialized subchannels (e.g. the bug detector cells and the moving-edge detectors, [3]. In rabbits too, specialization of ganglion cells in the discrimination of objects moving at different speeds lead to subtypes division thus showing at least two systems for motion description, [1], and the extensive study of the direction selective ganglion cells has yield a direct correlate between functional properties as the response to sinusoidal gratings and structural design, e.g. receptive field diameter [2].

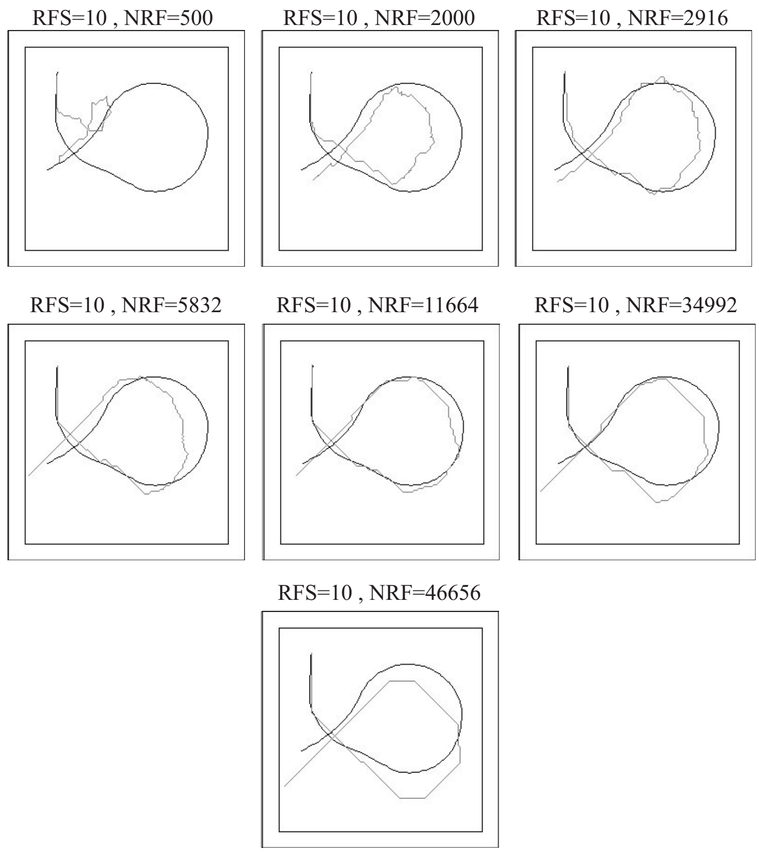

Figure 5: The effect of varying receptive field number with a fix RF size of 10 pixels. (NRF=Number of Receptive Fields randomly distributed over the input image). The solid darker line shows the real trajectory of the object; the lightest shows the calculated trajectory using data from speed estimation and direction selectivity cells.

Given a determined experimental species, network parameters like number of cells or receptive field size within each motion detection subchannel are constant, or at least no significant variations between individuals has been reported. In rabbit, for instance, 45% of all ganglion cells are specialized in movement detection and they have been divided in two specialized classes: one for fast moving land predators, the other for slow moving airborne predators, [1], while in frog, movement detection devotes at least 30 times more fibers (e.g. ganglion cells) than to any other image descriptor [3], one strictly focused in bug catching, the other in more general moving edges. Every channel or subchannel is, then, perfectly designed to detect whatever descriptor is tuned to. Moreover, in order to the system to work properly, a minimum number of cells is necessary within the channel and a distribution that covers the greater part of the retina is needed. But, what are the optimal figures for cell population? A very small number of cells may make the system inoperative, so one can expect that the more cells, the more accurate the information on moving objects will be. On the other hand an excess of detecting cells may have undesirable performance for the network, as it does happen in our computational model, where increasing the computing units from a certain point has a negative effect in estimating the position of the moving object. Increasing the number of receptive fields increases redundancy and in this case redundancy is rather a problem than a solution. Thus, too many specialized cells imply a higher computational cost for the retina and, which is worse, the possibility of being caught by predators due to the degrading performance of the positioning subsystem.

4 Conclusions Channel processing is a key computing mechanism both for natural and artificial visual systems: it keeps computational complexity within tractable limits, permits the extraction of several visual descriptors in parallel and allows reliability against computing unit (cells) failures. This channel division of the overall network, however, seems to work in a recursive way in the sense that each channel (e.g. motion) is usually divided in sub channels (e.g. low vs fast moving objects detection) permitting a finer analysis of the outside world. A system containing several subchannels might be more reliable and less costly (both in complexity and computational operations) than a single do-it-all channel, even when the number of processors of this last channel could be less than the total sum of them in the former scheme.

The developed example of an artificial, but biologically flavoured, visual system gives us a hint on how individual function and network processing influences each other. One unexpected characteristic is that of the degradation of results in trajectory estimation when the size of receptive fields or the number of receptive fields, i.e. of computing cells, increase. To get good results for a variety of moving object speeds we have to reach a compromise between size and number of cells (position of cells in the processing network is random). Both a lack of and an excess in the number of computing elements lead to a bad performing network in trajectory description tasks. An explanation of the effect resides in the fact that when the size or RF is big the direction selectivity mechanism is affected, thus misleading the overall system behaviour. Also, an excessive number of receptive fields increases redundancy in the system and in this sense, redundancy seems to be rather a problem than a solution. This can be used as an advantage, since loss of processors (reducing the number of receptive fields) need not dramatically affect the performance of the movement detection and analysis system provided there is a minimum overlap and the RF sizes are big enough to cope with a range of speeds. Computational cost, then, can be also controlled to a minimum and benefits from this property.

References

[1] Famiglietti, EV 2005. "Small tufted" ganglion cells and two visual systems for the detection of object motion in rabbit retina. Visual Neuroscience 22, 509–534.

[2] He S, Levick WR 2000. Spatial-temporal response characteristics of the ON-OFF direction selective ganglion cells in the rabbit retina. Neuroscience letters 285, 25–28.

[3] Lettvin J, Maturana A, McCulloch WS, Pitts W 1956. What the frog's eye tells the frog's brain. In WS McCulloch Embodiments of Mind, The MIT Press, Cambridge, Mass.

[4] Peichl L. 1989. Organization of the retina: structure/function relations and a species comparison of retinal ganglion cells. Fortschr Ophthalmol 86(1), 47–53. (In german).

[5] Quesada-Arencibia, A. 2001. Un sistema bioinspirado de análisis y seguimiento visual de movimiento. PhD Dissertation, Universidad de Las Palmas de Gran Canaria. (In spanish).

[6] Quesada-Arencibia, A., Moreno-Díaz jr, R., Alemán-Flores, M., 2001. Biologically Based CAST-Mechanism for Visual Motion Analysis. Springer Lecture Notes in Computer Science 2178, 316–327

[7] Rodríguez-Rodríguez, J.C., Quesada-Arencibia, A., Moreno-Díaz jr, R., Leibovic, K.N., 2003. On Parallel Channel Modeling of Retinal Processes. Springer Lecture Notes in Computer Science 2809, 471–481

Roberto Moreno-Díaz jr, Alexis Quesada-Arencibia, Gabriel de Blasio
Instituto Universitario de Ciencias y Tecnologías Cibernéticas, Universidad de Las Palmas de Gran Canaria
35017 Las Palmas, Spain
E-mail: {rmorenoj, aquesada, gdeblasio}@dis.ulpgc.es

José Carlos Rodríguez-Rodríguez,
Instituto Universitario de Ciencias y Tecnologías Cibernéticas, Universidad de Las Palmas de Gran Canaria
35017 Las Palmas, Spain
E-mail: jcarlos@ciber.ulpgc.es