A DEDICATED NEURAL NETWORK FOR VISUAL MOTION DETECTION

Emmanuel Hanssens, Michel Verleysen & Jean-Didier Legat

Université Catholique de Louvain
Laboratoire de Microélectronique
Louvain-La-Neuve, Belgium

***Abstract** : This paper presents a neural network for visual motion detection. The network uses a novel method to compute the velocity of an object, based on the knowledge of the motion history. This method solves the well-known aperture problem by giving the global velocity. The network has two layers of cells ; the first layer computes the local velocity directly from the image pixels, whereas the second layer can be seen as a decision layer that remembers the history of the motion to transform the outputs of the first layer into the global velocity. Simulations show the effectiveness of the network, and the importance of a parameter named remanence on the motion determination.*

***Keywords** : neural network, visual motion, global velocity, decision layer, lateral connections chains*

Résumé : Nous présentons ici un réseau de neurones destiné à la détection de mouvement. Ce réseau utilise une méthode originale pour calculer la vitesse d'un objet, basée sur la connaissance de l'histoire du mouvement. Cette méthode solutionne le problème bien connu d'ouverture en déterminant la vitesse globale. Le réseau comporte deux couches de cellules ; la première couche peut être vue comme locale directement à partir des points de l'image, tandis que la seconde couche peut être vue comme une couche de décision qui se rappelle l'histoire du mouvement afin de transformer les résultats de la première couche pour obtenir la vitesse globale. Des simulations ont mis en évidence l'efficacité du réseau, ainsi que l'importance d'un paramètre, appelé rémanence, sur la détermination du mouvement.

Mots clés : réseau de neurones, mouvement visuel, vitesse globale, couche de décision, chaîne de connexions latérales

---

Sponsored by the National Fund for Scientific Research (Belgium)

1. Introduction

From a human point of view, it has always been natural "to see". During his whole life, a human being sees what he looks at, without efforts. However, vision is one of the most complex tasks performed by humans: it requires many neural materials and interconnections implying highly complex organizations (Van Essen & Maunsell, 1983).

One of these tasks is motion detection (Allman, Miezin & McGuinness, 1985; Felleman & Van Essen, 1987). There are in the brain specialized structures aimed to detect and analyze motion, gathered in what is called the visual cortex. The basic feature of such networks is the high parallelism. It is also important to note that these structures resolve a fundamental visual processing problem known as the *aperture problem* (see below); therefore, it seems very attractive to take inspiration from these natural structures to mimic such an interesting feature.

These networks can be used in dynamic image segmentation and recognition. Relative motion allows mobile robot to navigate quickly and efficiently through the environment.

This paper presents an original network based on certain structures of the visual cortex and able to perform motion detection. We will show how the network behaves regarding the aperture problem, after explaining what this problem consists in. Other features of the network will be explored and some simulations will figure how it works.

2. The aperture problem

The aperture problem (Hildreth, 1983; Marr, 1982; Adelson & Movshon, 1982) arises from the fact that single photoreceptive cell cannot determine the actual motion of an object's edge. Since each cell's receptive field is sensitive to visual stimuli only within a spatially local region, or "aperture", the direction of an edge moving across this region is ambiguous.

Following Horn and Schunk (1981), we denote the image brightness at point (x,y) in the image plane at time t by I(x,y,t). When a pattern moves, the brightness of a particular point in the pattern is constant, so that:

$$\frac{dI}{dt} = 0$$

By differentiation, we obtain:

$$\frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t} = 0$$

where $V_x = \frac{dx}{dt}$, $V_y = \frac{dy}{dt}$, the velocity components, are the two unknowns of this simple linear equation.

This relation by itself is not sufficient to determine the velocity flow. It only defines a constraint line in velocity space that has the same orientation as does the edge of the moving pattern in physical space. This ambiguity is known as the aperture problem and is illustrated in figure 1.

Since each edge of a moving object generates a constraint line, the standard way to determine the actual velocity of an object is to compute the intersection of these lines. Figure 1 shows that two edges give a different constraint line in the velocity space (plane V_x - V_y). The intersection of these two lines indicates the velocity that satisfies both constraints and is then the actual velocity of the object. Nevertheless the establishment of these constraint lines requires the computation of the brightness partial derivatives with respect to time and axes. Furthermore, the intersection of several constraint lines is not always unique due to the approximations used for the derivatives. A refined technique must then be used to compute the velocity.

In this paper, we present a method that uses velocity informations at the same locations but at different times, instead of instantaneous velocity constraints at different locations, to compute the velocity field associated with a moving object.

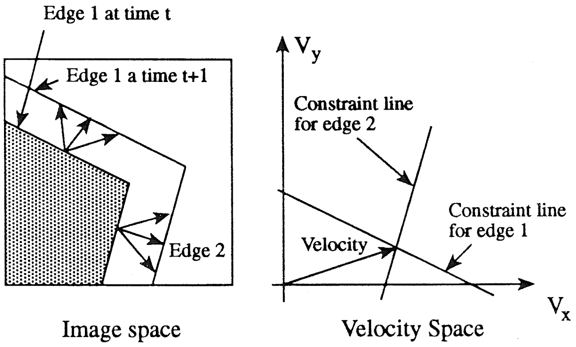


Figure 1. The aperture problem and velocity determination by intersection of the constraint lines.

3. A Motion Processing Network

As described above, the velocity can be estimated by the intersection of constraint lines corresponding to different edges of the moving pattern. By this way, we combine two (or more) measures at the same time but at different locations. There is another way for computing the velocities: instead of using velocities at different locations, we use velocities at different times, or, in other words, the time "history" of the motion. The reason for this is that the velocity of an object cannot change instantaneously. A network based on this principle will use previous computations of the velocity for determining the actual velocity. Such a network performs "tracking" of visual features as they move across the visual field. A simple network with two layers of cells will achieve this task, as shown in this section; the two layers are named L_1 (input layer) and L_2 (processing layer).

A similar network is described by Marshall (1989), but it uses for L_1 idealized cells analogous to the hypercomplex cells found in the visual cortex, i.e. cells sensitive to the orientation (Frégnac & Imbert, 1984), length and local direction of motion (Cremieux, Orban, Duysens & Amblard, 1987; Kennedy & Orban, 1983) of visual stimuli in their receptive field. We will use for L_1 simpler cells more adapted to the VLSI implementation of the network. Indeed, we do not need cells sensitive to the orientation and the length of an object's edge, but only to the local direction. These cells detect the local motion of the edge by comparing two neighboring pixels, one delayed with respect to the other. The pixels come from light photodetectors.

In the following, we will give a description of the topology of the network. In section 6, simulations of a simple example will show how it works.

As already said, the network contains two layers of cells (L_1 and L_2). The first layer (L_1) cells project excitatory connections forward to cells in L_2. Each L_1 cell projects to a cluster of neighboring L_2 cells (Figure 2). There are as many cells in a cluster as there are directions to detect.

As shown on figure 2, we choose four cells in each cluster; however, this network can be extended without any problem to other directions. This means that the network will have four preferred directions (North, West, South, East). This does not mean that the network will be unable to detect motion in other directions: we will see in next section how the network can deal with such other directions.

On each L_1 cell locations, the network detects the local direction of the motion, and then excites the L_2 cell corresponding to that direction in the cluster attached to the L_1 cell. Then, the layer L_2 determines whether the local direction given by the layer L_1 is compatible with the global direction of

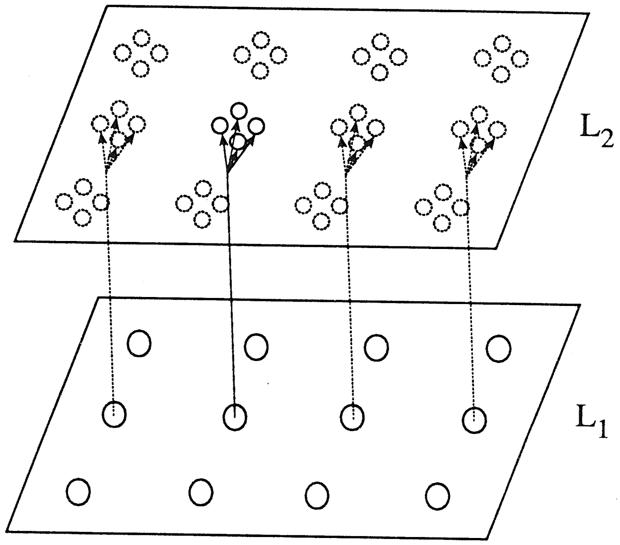


Figure 2. The excitatory connections from cells L_1 to a cluster of cells L_2.

the motion. This is done by means of the lateral connections in layer L_2: each L_2 cell in a cluster receives a strong excitatory input connection from an L_2 cell displaced spatially in one direction, and sends a strong excitatory output connection to another L_2 cell displaced in the opposite direction. For instance, if we denote by "North L_2 cell" the cell in a cluster that detects a motion in the North direction, each North L_2 cell receives a strong excitatory input connection from the "North L_2 cell" of the previous cluster in that direction (that is, the cluster located just at the South), and sends a strong excitatory output connection to the "North L_2 cell" of the next cluster in that direction (located at the North). Figure 3 depicts the embedding of an L_2 cluster in its lateral connection chains.

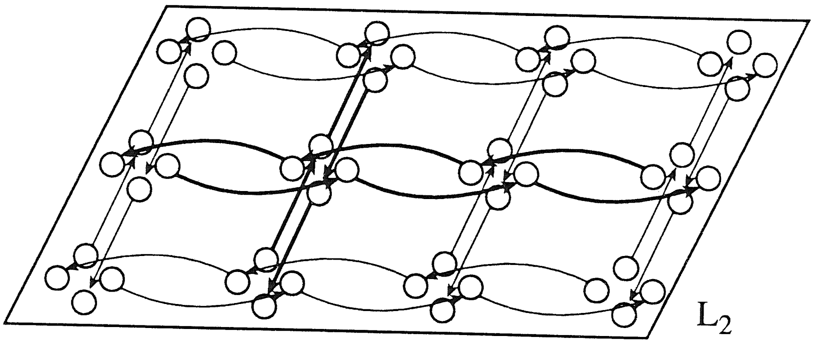


Figure 3. Lateral chains of excitatory connections. Each cell of a cluster participates in a different chain. The chains for one cluster are shown with bold arrows.

Moreover, these lateral connections possess a signal transmission latency. In other words, a signal emitted by one cell does not reach its destination cell until a prescribed time later. The timing of the lateral transmission latencies figures prominently in the operation of the network. It will determine a speed motion for which the detection will be optimum. At this speed, the distance between two photodetectors is covered during the latency. For speeds very different from that optimal speed, the motion will not be detected any more. This latency should then be adjusted to give the desired speed range. If the network has to detect several speeds at one time, then it should be provided with some other L_2 layers with different transmission latencies.

The final element in the structure of L_2 is the set of lateral inhibitory connections between cells in each cluster (Nabet & Pinter, 1991). Figure 4 shows these connections for one cluster. With these inhibitory connections, each cluster will indicate only one direction of motion, since only one cell in each cluster will be excited at once.

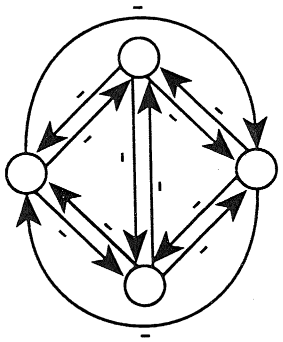


Figure 4. The inhibitory connections between the L_2 cells of a cluster.

The network described above is able to compute the global velocities of moving patterns. We will see in section 6 the simulation results (activity of L_2 cells) for a simple motion.

4. Motion in a non-preferred direction

When the motion direction of a pattern does not correspond with one of the preferred directions of the network — that is, in this paper, the four directions North, West, South and East — it is a little more complicated to see how the "tracking" will be achieved. Let's suppose, for instance, that we have an edge moving across the visual field in South-East direction, as shown on figure 5. At layer L_1, the motion that will be detected is the local motion of the edge. Obviously, the edge seems to move in the East direction when we look at a small part of it. This is exactly the phenomenon described in section 2, the aperture problem. The L_1 cells will fire the L_2 cells whose preferred direction is East (called "East L_2 cells" in the previous section) since it corresponds to the local direction of the motion. The global direction will be determined here by looking at the set of L_2 cells that fire. This set is surrounding the arrow drawn on the figure 6, showing the movement of the rightmost edge.

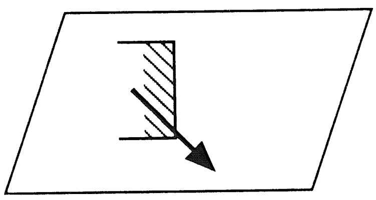


Figure 5. Example of non-preferred direction : South-East.

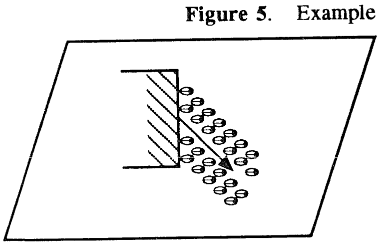


Figure 6. Movement of the right edge with the set of firing L_2 cells.

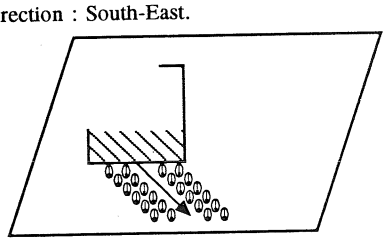


Figure 7. Movement of the bottom edge with the set of firing L_2 cells.

If we take a look closer to the bottom edge, we will observe that it is here the "South L_2 cells" that fire, as shown on figure 7. A set of firing L_2 cells can also be noticed, from which can be deduced the movement of the bottom edge (the arrow on the figure). Obviously, the two figures show the same movement, as the two edges are both parts of the same pattern. Hence, the two sets indicate the actual (or global) direction of motion, by their shape.

The fact is that the layer L_1 will detect the velocity component perpendicular to the moving edge (local motion). For this reason, the example shown here is one of the worst because the two edges are perpendicular to two preferred directions of the network. If this is not the case, the set of cells that will fire would be much more narrow. If necessary, a more accurate detection can be obtained by adding other L_2 cells in each cluster, with new connection chains in layer L_2. The L_2 structure with four cells per cluster leads to the optimal detection of four motion directions (North, West, South and East). For the detection of a combined direction such as South-East (the example given above), four new cells should be inserted in each cluster.

5. Cell activity remanence

When a motion is detected, the activity of some L_2 cells will increase to indicate the direction of the motion. As the motion is a time dependent event, the activity will also vary during the detection. We call remanence the time needed by the L_2 cell to return to its initial state, where the activity is considered to be zero (no motion detected or, equivalently, motion in all directions leading to the same activity for all L_2 cells of the cluster). This parameter is important and has to be related to the size of the moving patterns. Indeed, we have seen that the network is able to detect moving edges, but if the pattern is uniform in brightness, then no further motion will be detected until another edge passes across the L_1 cell. If the remanence is high, then the (detected) motion will "last" longer than with a low remanence. As a consequence, if we look at an instantaneous image, there will be motion at an edge and behind the edge. A set of L_2 cells will still fire (with decreasing activity) just where the edge passed some time ago. The higher the remanence, the larger the set will be. This effect allows to detect motion at a location where there are no more variations of the brightness (inside the pattern): there will be a larger part of the moving pattern for which motion is detected (depending on the size and the remanence).

The drawback of this effect is that motion will also be detected behind the extremity edge (edge that closes the pattern) where no real motion exists. This can become confusing if a second pattern passes before the activity of the cells has returned to "zero". For that reason, the remanence should not be set at a very high value. Figure 8 depicts the sets of L_2 cells firing because of the remanence.

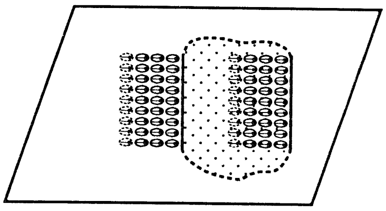


Figure 8. The remanence effect on the activity of the L_2 cells.

6. Simulations

In this section, we will show the results of simulations for a simple motion example. The pattern considered is a square (10 x 10 pixels) moving toward East on a uniform background (32 x 32 pixels) as shown on figure 9. The speed of the square is chosen to give the optimal detection (in other words, the square moves by one pixel at each image).

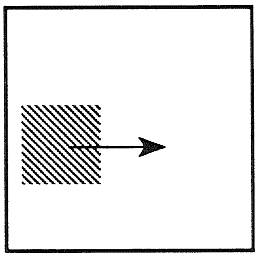
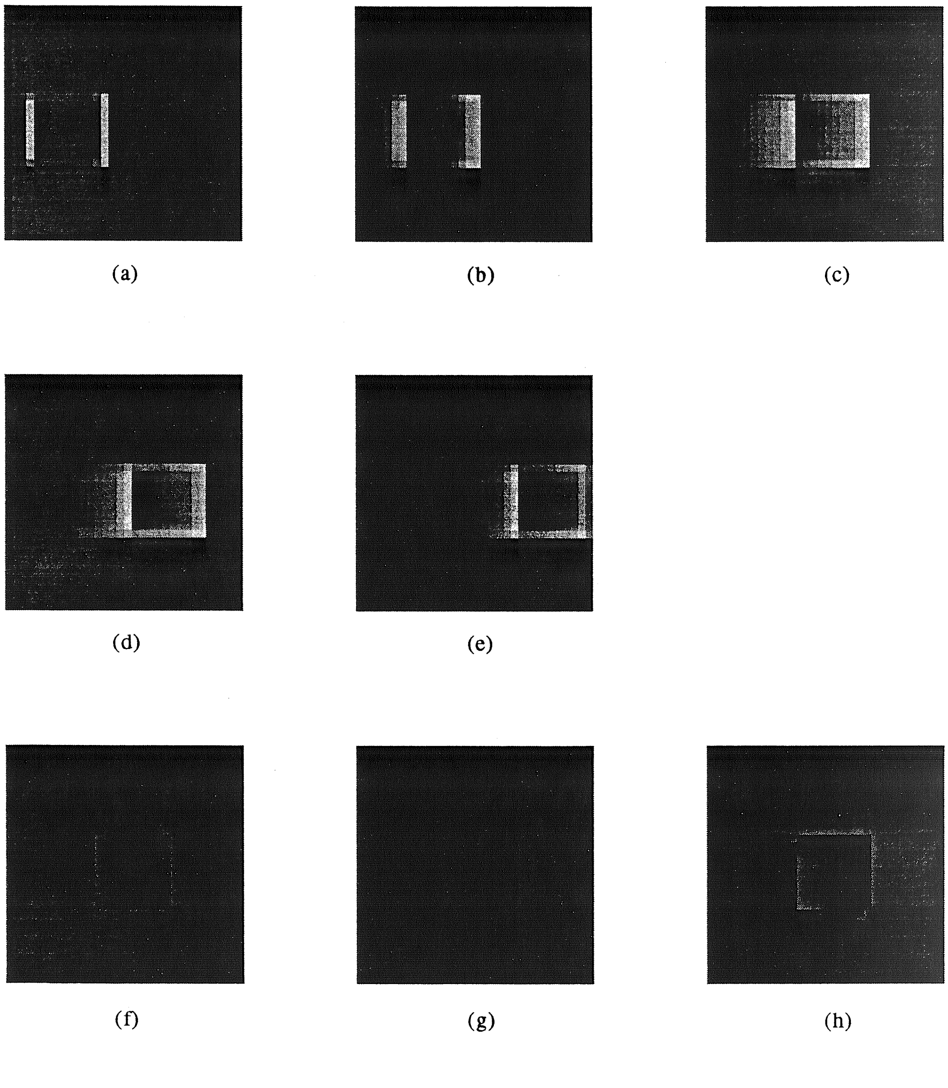


Figure 9. A simple motion example for the simulations.

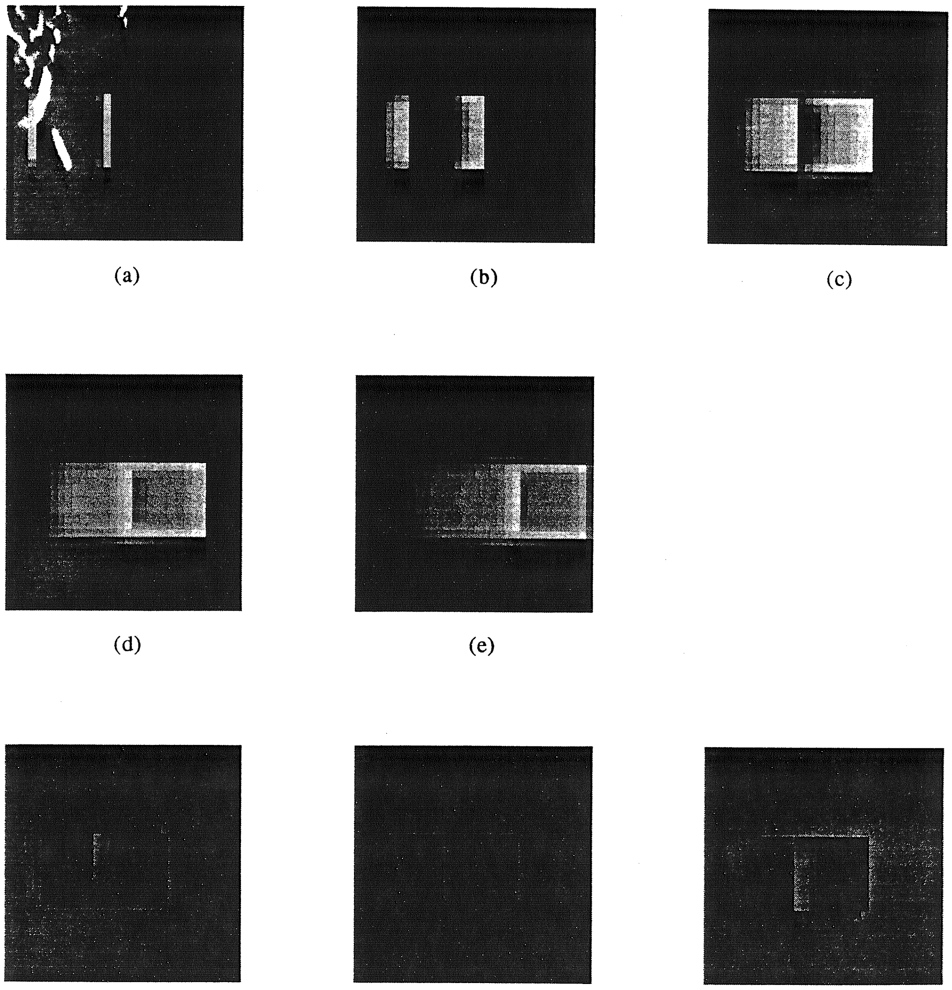


(a) (b) (c)

(d) (e)

(f) (g) (h)

Figure 10. Normal remanence. (a) - (e) East L_2 cells activities at t=2, t=5, t=10, t=15 and t=20. (f) North L_2 cells activities at t=10. (g) West L_2 cells activities at t=10 (h) South L_2 cells activities at t=10.



(a) (b) (c)

(d) (e)

(f) (g) (h)

Figure 11. High remanence. (a) - (e) East L_2 cells activities at t=2, t=5, t=10, t=15 and t=20. (f) North L_2 cells activities at t=10. (g) West L_2 cells activities at t=10. (h) South L_2 cells activities at t=10.

The pictures on figures 10 and 11 show the evolution of the activities of the L_2 cells, figure 10 for a normal remanence and figure 11 for a high remanence. A high activity is represented by a white pixel, "zero" activity by dark gray, while a black pixel means no motion in that direction but well in another direction.

Pictures (a) to (e) correspond to the "East L_2 cells" at time t=2, 5, 10, 15 and 20 (time t=1 is for the first image when the motion begins, t being incremented at each new image). Pictures (f), (g) and (h) show the activities of, respectively, the "North L_2 cells", the "West L_2 cells" and the "South L_2 cells" at time t=10.

The pictures show that the detection is maximum at the vertical edges, as expected since they are perpendicular to the direction of motion. At the beginning (first picture), the "history" of the motion is short; therefore, the sets of firing L_2 cells are small, showing motion (toward East) only at the two edges. After a while, the sets grow because of the remanence (if the remanence were zero, we would only notice motion at the two edges, like on the first picture), and the differences between the two simulations become clear.

It is also interesting to observe what happens at the horizontal edges. As there is no motion perpendicular to these edges, the "North-" and the "South L_2 cells" activities are very low, indicating a motion in one of the two other directions. Indeed, there is a motion toward East, but it is only the history of the motion that raises the "East L_2 cells" activities (and lowers the "West L_2 cells" activities) at this location. From layer L_1, we would see that there is no local motion detected at these edges.

7. Conclusion

Although motion detection is a very complex task, which requires many neurons in the brain, it has been shown that a simple two-layers network can mimic some of the features of the visual cortex, especially the gathering of local direction computations into a decision layer (L_2) determining the global direction of motion. There is no problem any more with the aperture; the reason is that the network "remembers" the history of the motion to "track" it across the visual field. The network has been successfully simulated on a standard Sparcstation by using some synthetic images.

Moreover, the network can be adapted to the size of the object by adjusting the remanence (within certain limits). It has been shown how this factor influences the evolution of the L_2 cells activities.

On the other hand, the outputs of the network (activities of L_2 cells) can be used for further image processing. Indeed, it is possible to build other layers of processing units on the top of L_2 to analyze special features of the motion. Speed selection, for instance, or detection of moving objects with a certain size,...These kinds of process can be easily achieved by using the network discussed in this paper.

References

Adelson, E. H., & Movshon, J. A. (1982). Phenomenal coherence of moving visual patterns. *Nature*, **300**, 523-525.

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (MT). *Perception*, **14**, 105-126.

Cremieux, J., Orban, G. A., Duysens, J., & Amblard, B. (1987). Response properties of area 17 neurons in cats reared in stroboscopic illumination. *Journal of Neurophysiology*, **54**, 211-222.

Felleman, D. J., & Van Essen, D. C. (1987). Receptive field properties of neurons in area V3 of macaque monkey extrastriate cortex. *Journal of Neurophysiology*, **57**, 889-920.

Frégnac, Y., & Imbert, M. (1984). Development of neuronal selectivity in primary visual cortex of cat. *Physiological Reviews*, **64**, 325-434.

Hildreth, E. C. (1983). Computing the velocity field along contours. *Proceedings of the ACM SIGGRAPH / SIGART Interdisciplinary Workshop on Motion* (pp. 26-32). Association for Computing Machinery.

Kennedy, H., & Orban, G. A. (1983). Response properties of visual cortical neurons in cats reared in stroboscopic illumination. *Journal of Neurophysiology*, **49**, 1071-1095.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco: W. H. Freeman and Company.

Marshall, J. A. (1990). Self-Organizing Neural Network for Perception of Visual Motion. *Neural Networks*, **3**, 45-74.

Nabet, B., & Pinter, R. B. (1991). *Sensory Neural Networks: lateral inhibition.* Boca Raton: CRC Press, Inc.

Van Essen, D. C. , & Maunsell, J. H. R. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, **6**, 370-375.