Neural competitive structures for segmentation based on motion features

Javier Díaz[1], Sonia Mota[1], Eduardo Ros[1] and Guillermo Botella[1]

[1] Departamento de Arquitectura y Tecnología de Computadores, E.T.S.I. Informática, Universidad de Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
{jdiaz, smota, eros, gbotella}@atc.ugr.es

Abstract. Simoncelli & Heeger studied how the motion is processed in humans (V1 and MT areas) and proposed a model based on neural populations that extract the local motion structure through local competition of MT like cells. In this paper we present a neural structure that works as dynamic filter on the top of this MT layer and can take advantage of the neural population coding that is supposed to be present in the MT cortical processing areas. The test bed application addressed in this work is an automatic watch up system for the rear-view mirror blind spot. The segmentation of overtaking cars in this scenario can take full advantage of the motion structure of the visual field provided that the ego-motion of the host car induces a global motion pattern whereas an overtaking car produces a motion pattern highly contrasted with this global ego-motion field.

1 Introduction

The work presented in this paper has been developed in the framework of ECOVISION European Research project [1]. One of the objectives of ECOVISION consortium is the development of a pre-cognitive visual model that can be useful for real world problems. In particular, the rear-view mirror blind spot monitor is presented as a feasible problem in which motion processing can provide useful information for the overtaking car segmentation.

One of the motion processing models studied in the first stage of the project is the Simoncelli and Heeger model (S&H) that constitutes a model with strong neurophysiological bases. The S&H work proposes a model of how the cortical areas (V1 and MT cells) can extract the motion structure through neural local computation and competition [2,3]. The output layer uses neural population coding, which can be seen as an inefficient code but represents and advantage if the post-processing is done through neural computation as presented in this paper. This is not the case if the optic flow extraction is done through more mathematical based algorithms.

The MT cells specialization on specific movement direction and velocity modules enables a connectivity that can embody perspective deformation correction and rigid body motion enhancement. This can be achieved by a collective-competitive connectivity pattern as described in section 3. Motion processing is normally very noisy and needs of further post-processing before addressing image segmentation. In this paper we describe how a simple connectivity pattern can be of interest for neural computation of noisy motion information. This connection pattern forces individual cells to behave as dynamic filters that are sensitive to more reliable movement features than simple spatio-temporal correlations.

This post-processing layer is composed by cells that collect the cell activity from MT cells sensitive to similar motion primitives, and compete between themselves to impose a movement feature locally in each visual field point. Furthermore, we also describe how this can enhance the rigid body motion segmentation capabilities of the processing layer by connecting MT cells of local neighbourhoods to facilitate the detection of movement features of rigid bodies through the visual field.

The application of the neural processing strategy presented in this paper in real world problems is also addressed. In particular, promising results have been obtained for the segmentation of overtaking cars in the rear-view mirror blind spot. This application is currently being addressed by many application driven research groups [4]. Besides, in this application the motion processing plays an important role, since an overtaking car exhibits a forward motion pattern clearly contrasted against the global backward motion pattern observed in the rear-mirror due to the ego-motion of the host car.

2 Bio-inspired model for computing optical flow

The Simoncelli & Heeger model [2,3], consists of two primary stages corresponding to cortical areas V1 and MT. The computation form is highly parallel and regular.

A linear model is used for V1 simple cells. This explains simple cell selectivity for stimulus orientation and spatial frequency, and the cells respond to both opposite polarities contrast stimulus.

V1 complex cells sum weighted simple cell afferents distributed over a local spatial region, each of them having the same space- time orientation and phase. Their receptive fields are modeled using only edge detectors. Each V1 neuron squared and normalizes its inputs, and V1 outputs with the same space-time orientation are spatially combined using positive weights to get V1 complex cells receptive field.

MT cells are modeled combining the outputs of a set of direction-selective V1 complex cells, whose preferred space-time orientations are consistent with the MT cells characteristic velocity. The mechanism for velocity selectivity can be described in the spatio-temporal domain easily. The power-spectrum of translational pattern lies on a plane, and the tilt of the plane depends on the velocity; in this way a MT cell detects the tilted plane with maximum response [5] and different combinations of V1 cells can be used to get the MT receptive field [6,7]. Finally, a Winner Takes All configuration among the MT population selects only the MT cells with higher input, i.e., the one that best matches the local motion pattern.

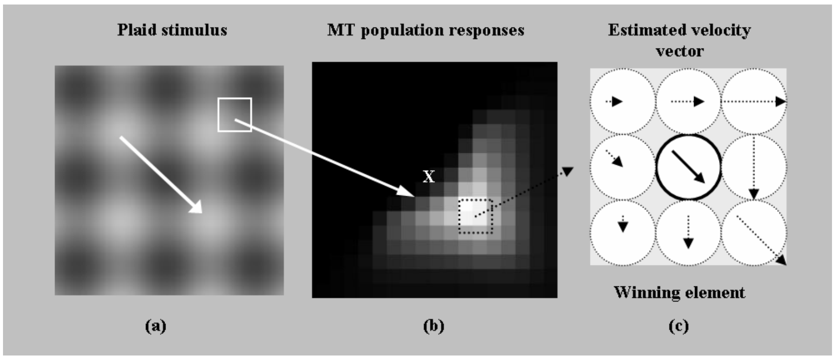


Fig. 1. Population response example. The result of a plaid stimulus composed by a sinusoidal grating moving rightwards and another moving downwards is a moving pattern toward the right-bottom corner (a). A set of MT neurons responses are codified using grey levels. The relative position of the winner element with respect to the center of the population, showing the velocity module and direction. Maximum responses are given at the best tuned MT neuron for that stimulus, but MT cells tuned at near velocities are not zero (b). Finally, the winner element indicate the estimate velocity (c).

The implementation of S&H model, showed in fig. 2, can be summarized in 4 steps:

1. Computing local contrast stimulus.
2. Modeling V1 simple and complex neurons, using spatio-temporal third Gaussian derivatives and spatial pooling.
3. Modeling MT neurons summing the weighted responses of V1 cells which lie on its characteristic plane.
4. Computing winner element for each pixel in the visual field

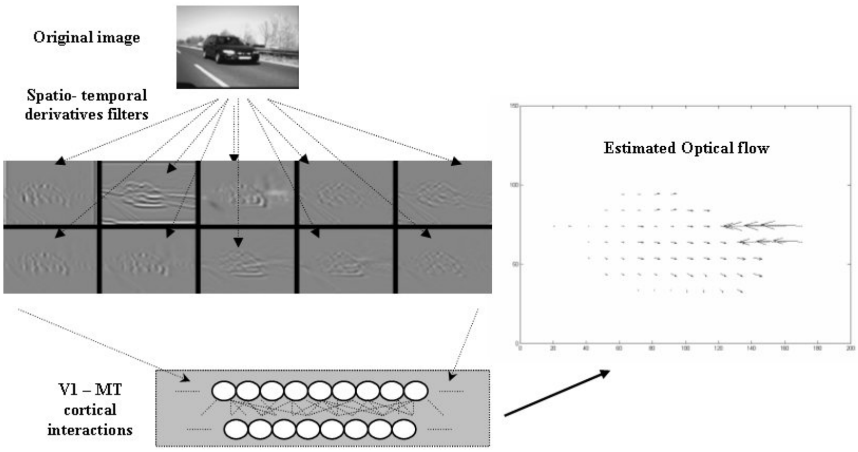


Fig. 2. S&H Model. An overtaking car sequence is used to evaluate the model. The basic set of third Gaussian derivatives are formed by G_{xxx}, G_{yyy}, G_{ttt}, G_{xxy}, G_{xyy}, G_{xxt}, G_{xtt}, G_{yyt}, G_{ytt} and G_{xyt}. This pre-filtered images are combined to get V1 spatio-temporal orientation cells and their combinations give us the MT receptive field. Finally, for each pixel the winner take all neuron is considered the correct vector velocity value.

The basic model used to describe V1 simple cell receptive field is an edge detector that uses spatio-temporal third Gaussian derivatives [8]. For our considerations we do not take into account other possible receptive fields such as bars detectors or DOG's like. Our implementation uses a basic spatio-temporal set of 40 filters with only one spatial scale to interpolate the MT characteristic orientation because biological systems have only a limited set of V1 orientations [9]. As other energy models [10], the contrast problem is hard to solve and some kind of normalization techniques [11] are used to minimize this effect. Finally, the MT neurons sum the weighted contribution of V1 cells near its preferred velocity.

One model limitation is the detection of second order motion. This kind of motion has power spectrum that lies out of the origin so the proposed filters can not detect it. Some modifications could be added to detect second order motion [12], but for the addressed application, in which we are just interested in translational motions, this is not necessary.

3 Dynamic Filters

The S&H layer is connected to a new neural layer that we have called Collector Layer (CL). The synaptic connection is done through a converging many-to-one excitatory pattern.

CL cells work as dynamic filters to segment the overtaking vehicle. Considering the kind of input information, the CL cells select only the features necessary for the segmentation process. First of all, because of the application addressed is focussed in discriminating between leftward (ego-motion) and rightward (overtaking vehicle) moving points, only the cortical S&H neurons that match these directions (leftward by [-135,215] degrees and rightward by [-45,45] degrees) are connected to the Collector Layer. In the other hand, the configuration of the collector layer neurons embodies different aspects about the character of a moving object that are important for the segmentation task: rigid body motion and scene perspective motion pattern deformations.

All the points in a rigid body move at a similar speed and in the same direction. The presence of detached points, belonging to a rigid body, that move in opposite direction with respect the majority of the points at the rigid body is considered noise. In addition, it is expected all points in a rigid body to be placed in the same neighbourhood of the image.

The motion pattern deformation due to the perspective from the rear-view mirror can be summarized as follows. A moving object (overtaking car) with a constant speed is expected to move slowly when it is localized in the very left side of the image (far away) and its speed will increase as it moves (overtakes) rightwards through the visual field (closer position). Considering this, forward sensitive neurons on the left side of the visual field tuned to higher speeds are less frequent than neurons placed on the right side and vice versa.

The S&H outputs stimulate the CL composed by self-competing collector neurons. Every collector neuron is tuned to a neighbourhood of a characteristic module velocity (by converging connections), in a fix direction. Therefore, every CL neuron integrates the activity of all the outputs from a 5x5 neighbourhood in the S&H layer (upper framed zones with continuous line in Fig. 3) sensitive to the same motion direction and velocity module.

The CL has the configuration of a self-competitive layer, so the collector neuron that receives the maximum stimulus in its spatial influence area (bottom framed zone with continuous line in Fig. 3) inhibits the others and dominates in its local area (Winner Takes All). This helps to detect rigid body motion; in other words, we can find rigid bodies in areas where there are winner collector neurons because they receive the input excitation of a group of MT neurons that are sensitive to the similar velocities (in module and direction) and are positioned in the same spatial area.

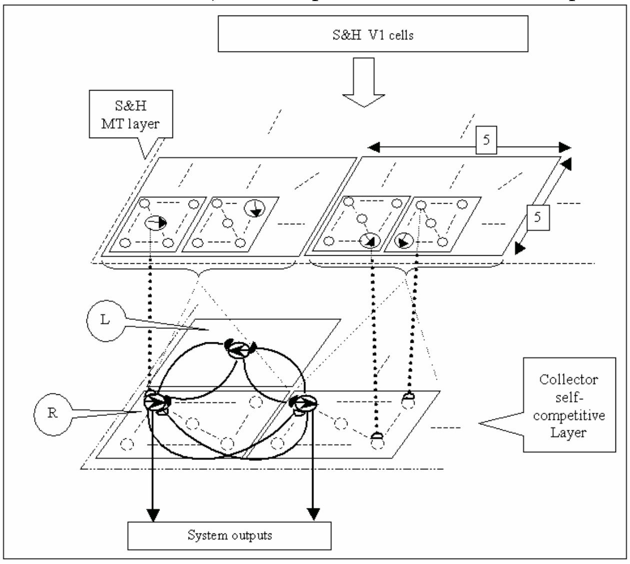


Fig. 3. Every collector makes spatial integration of the velocities (in module and direction) among a population of 5x5 S&H excitatory output neurons. The CL has the structure of a Winner-Takes-All layer, so there is just one collector neuron winner in a neighbourhood. A winner node receives synapses from other winners in an influence area. A group of neurons that detect the same direction support each other (white exciting synapsis), but if there is a neuron detecting an opposite direction it is inhibited by the synapses coming from other nodes corresponding to the same spatial neighbourhood (black inhibiting synapsis). The upper figure shows the synapses among three winner collector neurons. Two neurons detect rightward motion direction (R) and the other detects leftward motion detection (L). The last one is inhibited by the other nodes (local majority).

In the other hand, the winner neurons in an influence area at CL (bottom in Fig. 3) can interact with other winner neurons from other influence areas in their neighbourhood. This interaction has inhibitory or excitatory character, facilitating the domination of large features and inhibiting those winner neurons whose motion direction is different with respect to the majority of the surrounding winner nodes. In this way, the output response of this filtering neural layer (CL) will be non-zero in areas where there are winner collector neurons non-inhibited by others winner nodes.

The CL neurons are characterized by a time constant that takes into account how the stimulus drives the onset and offset of the elements of this layer. If we choose this time constant to be long, that means that more input frames with a lasting motion pattern are needed to excite a neuron and make it dominate against previous perceived patterns.

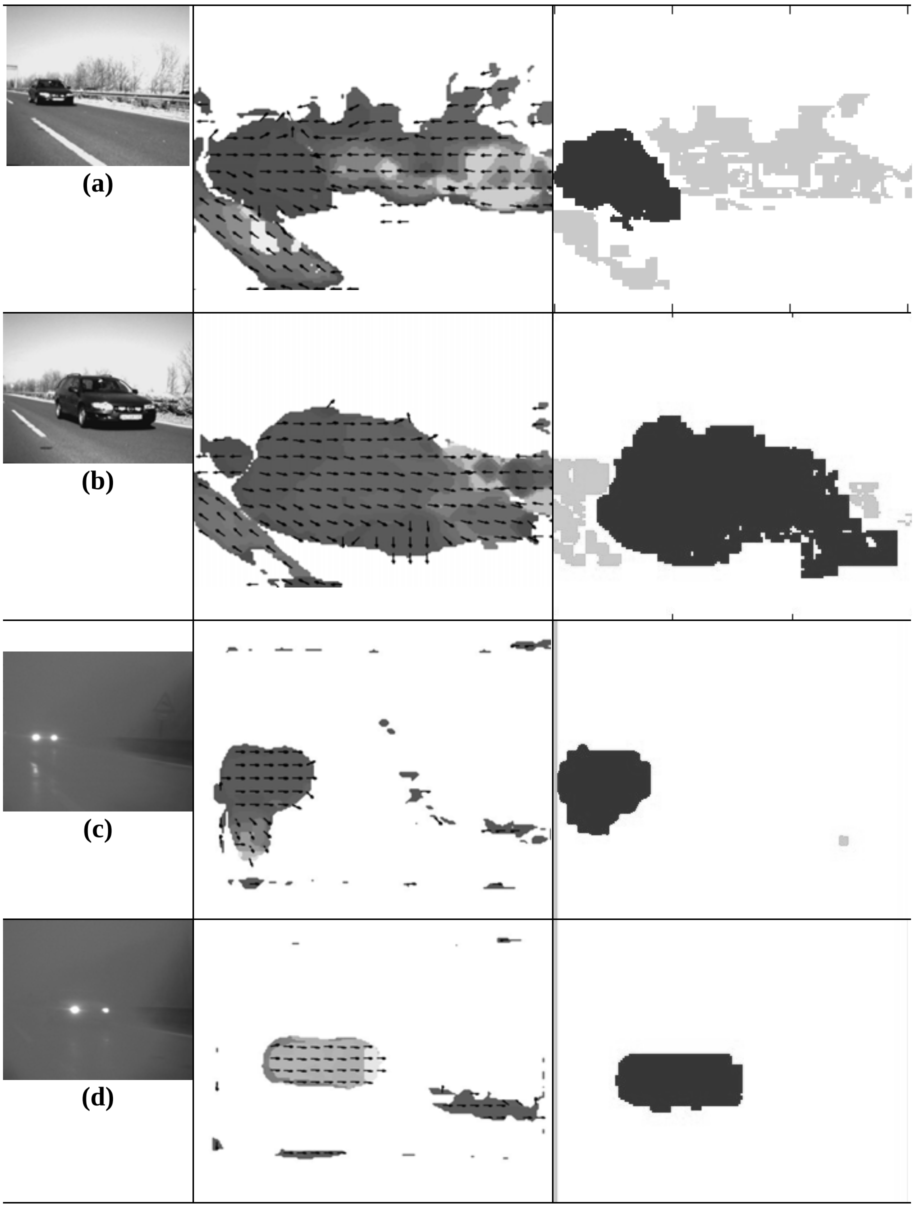
The distribution of the specialised collector neurons is non-uniform, i.e. the number of collector cells tuned at low velocities is higher in the left hand side of the layer than the number of collector neurons tuned at high velocities in the same place; and the opposite occurs in the right hand side of the layer. In this way, we facilitate the detection of slow movements in the left side of the visual field and rapid movements on the right side; this is used to correct the perspective deformation.

4 Results

We show the neural segmentation results of the previously described system applied to real overtaking car sequences. Our neural layers are capable of segmenting rigid objects that are moving in opposite horizontal directions.

Fig 4 (a,b), shows overtaking car sequence with dark car in a shinning day recorded with a conventional CCD camera. Left column are the original images of the overtaking sequences in two different stages, in the middle column the S&H extracted optical flow is shown. The grey scale indicates the velocity module and arrows show only the motion direction (all arrows have the same length). The right column shows the dynamic filters outputs. The segmented overtaking car is drawn using dark colour (rightward motion) and the background, moving thought the opposite direction, use bright colour. The collector layer receptive fields are sensitive only to synapses belonging to MT neurons tuned in a cone of velocities directions. For example, the bottom car optical flow in Fig 4.b indicate motion out of this allow velocities cone, so the collector layer output discards these components.

Other results are shown in Fig 4 (c.d). An overtaking car sequence in a foggy and rainy day recorded with high dynamic range camera [13].The weather produces a noise sequence with low contrast, therefore a special camera is necessary for this situation. Consequently, the extracted optical flow is worst than the one of the previous sequence. Other effects that contribute to get worst car segmentations are light reflection on the road and low contrast that produces more filter blurring. Another problem factor is the colour-depth. The high dynamic range camera uses 32 bits precision and we have used only 8 bits depth. This effect produces a noisy pattern that affects mainly the right side of the image and gives us wrong optical flow estimation in this area. But, it is clear the advantage of using dynamic filter outputs to segment correctly the overtaking car.



(a) (b) (c) (d)

Fig. 4. Overtaking car sequence in a shinning day (a,b) and in a foggy and rainy day (c,d).

5 Conclusions

This paper describes a bio-inspired system viability to segment objects using optical flow extracted from motion information. Motion information from V1 and MT layers is filtered by post processing layer that works as dynamic filters. The correction pattern from S&H MT cells this collector layer can embody aspect that facilitates the segmentation of moving rigid bodies and can also partially correct the deformation of the visual field due to the perspective of the rear view mirror.

The neural system proposed is highly parallel. It is a self-competitive neural mechanism for feature selection. All this enhances the capability of segment the rigid bodies from noisy environments.

Acknowledgements

This work has been supported by the V EU research framework funds through the European Projects ECOVISION (IST-2001-32114) [1] and CORTIVIS (QLK6-CT-2001-00279). We would like also to acknowledge Hella [14] for providing us with the overtaking sequences.

References

1. ECOVISION, http://www.pspc.dibe.unige.it/~ecovision/
2. Simoncelli Eero P, Heeger David J. (1998): A model of Neuronal Responses in Visual Area MT. Vision Research, 38(5):743-761.
3. Simoncelli, E. P.(1993). Distributed Analysis and Representation of Visual Motion. PhD thesis, Massachusetts Institute of Technology, Dept of E. Eng. Comp. Sci., Cambridge,MA.
4. U. Franke, D. Gavrila, A. Gern, S. Görzig, R. Janssen, F. Paetzold and C. Wöhler: "From door to door- Principles and Application on Computer Vision for driver assistant systems" in Intelligent Vehicle Technologies: Theory and Applications, Arnold 2000.
5. Watson, A.B, Ahumada, A.J. (1983). A look at motion in the frequency domain. In Tsosos, J.K., editor, Motion: Perception and representation, pages1–10. NewYork.
6. Grzywacz, N. M., Yuille, A. L. (1990). A model for the estimate of local image velocity by cells in the visual cortex. Proceeding so fthe Royal Society of London A,239,129–161.
7. Heeger, D.J. (1987). Model for the extraction of image of Image flow. Journal of the Optical Society of America A, 4, 1455–1471.
8. Freeman, W. T, Adelson, E.H. (1991). The design and use of steerable filters. IEEE Pattern Analysis and Machine Intelligence,13,891–906.
9. Hubel, D.H. y Wiesel, T.N.(1962). Receptive fields, binocular interactions and functional architecture in the Cat's Visual Cortex. J. Physiology, 160, pp. 106-154.
10. Adelson, E. H, Bergen, J.R. (1986).The extraction of spatio-temporal energy in human and machine vision. In Proc. of IEEE Workshop on Motion, pp. 151–156.
11. Heeger , D. J. (1992). Normalization of cell responses in cat striate cortex. Visual Neuroscience, 9, 181–198.
12. Sperling G., Chubb, C., Solomon, J.A., and Lu, Z-L (1994). Fullwave and halfwave proccesses in second order motion and tecture. Higher-order processing in the visual system. Chichester, UK: Wiley, pp. 287-303.
13. IMS-CHIPS: http://www.ims-chips.de/home.php3?id=e0841.
14. Dept. of predevelopment EE-11, Hella KG Hueck & Co., Germany, www.hella.de