ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

A neural model with feedback for robust disambiguation of motion

Mauricio Cerda and Bernard Girau

LORIA - INRIA Nancy Grand Est - Cortex team
B.P. 239, campus scientifique - 54506, Vandœuvre-lès-Nancy - France
{cerdavim,girau}@loria.fr

Abstract. The aperture problem is a direct consequence of any local detection in the visual perception of motion. It results in ambiguous responses of the local motion detectors. Biological systems, such as the brain of different mammals, are able to disambiguate motion detection. Such disambiguation is usually seen as a possible result of a pyramidal feedforward processing with growing receptive fields, but this approach is not able to detect motion in a simultaneously unambiguous and precise way. In this work we define a neural model of motion disambiguation that achieves both criteria, mainly with the help of excitatory feedback. Our model mostly differs from previous ones by incorporating lateral inhibition. Its main advantages are: tolerance to noise and stability. We perform tests on synthetic image sequences that show the effectiveness of our approach.

1 Introduction

Visual perception of motion is a major challenge in machine perception research, since it constitutes an important step in a wide variety of tasks such as path-finding, estimation of time to collision, perception of gestures, movement control, etc. In [1], we have developed a bio-inspired neural architecture that computes the optical flow and tracks one or several moving objects in a visual scene. Our model massively uses bio-inspired inhibitory/excitatory mechanisms that induce local competitions between antagonistic movements so as to improve the local coherence of motion perception. This work still faces many concrete difficulties, such as specular effects, shadowing, texturing, occlusion and the well-known aperture problem, responsible for ambiguous local detection of motion. In this paper, we address this latter problem.

More precisely, we define a neural model able to disambiguate motion perception for objects of different sizes, while maintaining the spatial precision achieved by local motion detectors. Instead of using several layers in a feedforward architecture, for which precision lies in the early layers while disambiguation may only be performed in the high-level layers, we combine some feedback and competition principles within the neural layers of [1] that are inspired by visual cortical areas V1 and MT (middle temporal). This results in a precise and unambiguous motion perception that proves very tolerant to noise.

In section 2 we give an overview of aperture problem and current solutions. Our model is presented in section 3, with corresponding results in section 4.

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

2 Aperture problem and disambiguation

In the context of optical flow extraction, many approaches have already been developed to solve the aperture problem. In this section, we are interested in the main existing models that are based on correlation and spatio-temporal energy filtering, as mentioned in [2]. The main reason for this is that these models better fit the current knowledge of how biological systems perform motion detection and integration. For example visual perception is commonly considered as a multi-stage process [3, 4]. We start with a short presentation of the well-known aperture problem, that causes the ambiguous local perception of motion.

2.1 Aperture Problem

The aperture problem always appears when the optical flow or image flow is estimated by means of local detectors. As illustrated in Fig. 1(a), only the motion component that is orthogonal to the local edge can be detected when looking into a small aperture. One important observation is that even though a local detector is inherently ambiguous, it is activated by local movements within a limited range (see Fig. 1(b)), so that several detectors may be combined to give unambiguous responses. This combination is called IOC (intersection of constraints, [5]). Another important observation is that local detectors are able to perform unambiguous detection for special features such as corners (Fig. 1(a)).

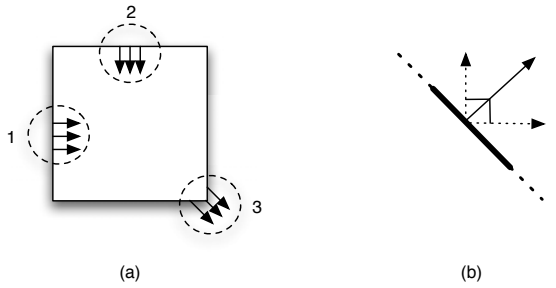

Fig. 1: A square diagonally moving (a), where circles represent the perceptive visual field of local motion detectors, and the arrows the detected motion. The local detectors give a range of possible motion of $\pm\frac{\pi}{2}$ around the direction that is, perpendicular to edge orientation (b).

2.2 Mechanisms of disambiguation

As it has been showed by [6], local motion detectors that use correlations (or region match) are equivalent to energy based filters (sensitive to a certain plane in the frequency domain), at least for the most common correlation schemes such as Reichardt Detectors (RD, [7]). We describe here the main strategies proposed

by other authors to integrate the local detection performed by either correlation or energy based motion filters.

2.2.1 Decoding the filtering output

In [8], Heeger uses as initial processing 3D Gabor energy filters, set according to 12 different spatio-temporal orientations that define a profile of activation for each given location. Heeger proposes a local normalization to deal with contrast problems. Then he considers the filter outputs as a code for the searched velocity, that he tries to retrieve by means of a least square technique. This approach finally reduces to finding the maximum value within a set of quadratic responses that estimate how much the observed filters fit predefined velocities.

In this approach, Heeger addresses three main ideas: 1) velocity selective filters require the combination of several frequency-selective filters, 2) locally true motion is encoded as the maximum response over the different velocities and 3) the importance of normalization to avoid that local changes in contrast are confused with actual differences in the responses of filtering. The main drawback of this technique is that it considers a fixed size of receptive fields. A test case as the one presented in Fig. 1(a) will not be completely solved: since the algorithm uses the information available at most at a distance equal to the size of the local filtering, an aperture problem that requires larger spatial integration will be only partially solved.

2.2.2 Interpolation

In [3], Simoncelli et al defines a more bio-inspired two-layer model of V1 and MT. Instead of explicitely computing the energy based filters, a series of oriented third derivatives of Gaussian are computed in V1. The way to get a velocity ($\vec{v}$) selective neuron in MT is to linearly combine four different spatio-temporal frequency selective filters whose response lays in the same plane in the frequency domain, forming a ring. Since these four filters are not necessarily represented in V1, they are computed as interpolations of V1 filters.

The main advantage of this model is that the velocity selectivity appears as much higher since it does not depend directly on the filtering stage. Additionally the two-level computation allows a much larger spatial integration. The main drawback is that the neighbourhood to perform this integration is still fixed. Any aperture problem that requires larger spatial integrations will eventually fails.

2.2.3 MT-V1 feedback

Bayerl et al [9] propose a mechanism based on the work of [3], in the sense that it has two sequential levels. The first level directly performs a locally normalized mechanism that is velocity selective (the first level of [3] is not immediately velocity selective, but rather frequency selective). The second level is also velocity selective and normalized. The main advantage is the introduction of feedback,

that appears to solve the aperture problems as long as two conditions are fulfilled: at least the corners of the moving object are contained in the image, and enough recurrent iterations are performed. Drawbacks of this approach are: 1) in the first level the normalization eliminates spatial differences in the responses of the filtering and 2) it does not take opposite orientations into account. The resulting lack of information use may result in a low tolerance to noise.

3 The Model

Our model is inspired by [9], trying to solve its identified problems and to provide both spatial precision and unambiguous motion perception. Our main contribution is to show that the principles of [9] may be improved by a competition that allows to locally maintain the preeminence of the most activated filters while strengthening the detectors that correspond to the true velocity. Thanks to this principle, a coherent and precise motion perception spreads along the edges of the moving object, with a high tolerance to noise thanks to the bio-inspired competition of [1].

The presented model is a sequence of two neural layers Ω_1 and Ω_2 (see Fig. 2(a)), where the outputs of the first one are the inputs for the second one. In terms of the total number of neurons: $|\Omega_1| \geq |\Omega_2|$. Normalization is achieved by shunting inhibition as proposed by [3]. The difference between both levels is that the second one performs a local competition to determine the predominant velocity as we proposed in [1], and the first one receives feedback from the second layer as proposed by [9].

Each layer contains neurons that are associated to the components of vector $\vec{x} = (x, y, \theta, v)$ that represent local motion detector parameters: a value of $\vec{x} = (2, 0, \frac{\pi}{2}, 1)$ in the first layer represents a local detector centered at spatial coordinates $(2, 0)$ specialize to detect motion at a speed of 1 pixel per frame to the right. Value $u(\vec{x}, t)$ (resp. $v(\vec{x}, t)$) stands for the potential of the neuron associated to $\vec{x}$ in layer Ω_1 (resp. Ω_2) at time t. At time t, several iterations are performed indicated with n. The update state equation for the first level (Ω_1) is defined as:

$$u_1^{n+1}(\vec{x}, t) = u_1^n(\vec{x}, t)\,(1 + cv^n(\vec{x}, t)) \quad (1)$$

$$u_2^{n+1}(\vec{x}, t) = \frac{u_1^{n+1}(\vec{x}, t)}{k + \sum_{a(\Omega_1)} u_2^n(\vec{x}', t) d\vec{x}'} \quad (2)$$

where $I(\vec{x}, t)$ is the input of the system and gives the initial input $u_1^0(\vec{x}, t) = I(\vec{x}, t)$, c is a control term for the level of feedback, and a represents adding in the same spatial location for all detectable speeds. The equation for the second level (Ω_2) then is:

$$v^{n+1}(\vec{x}, t) = \sum_{b(\Omega_1)} u_2^{n+1}(\vec{x}', t) s(\vec{x}', \vec{x}) d\vec{x}' \quad (3)$$

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

where $s(\vec{a},\vec{b})$ is the weight function that integrates different speeds detectors and b represents adding for all possible speed in the close neighborhood. To use this model, we have to define $s(\vec{a},\vec{b})$. Based on [9] and the competition of [1], we propose to use:

$$s(\vec{a},\vec{b}) \quad = \quad \exp\left(-\frac{(a_x-b_x)^2+(a_y-b_y)^2}{\sigma^2}\right) sgn(a_v b_v)\cos(a_\theta - b_\theta) \qquad (4)$$

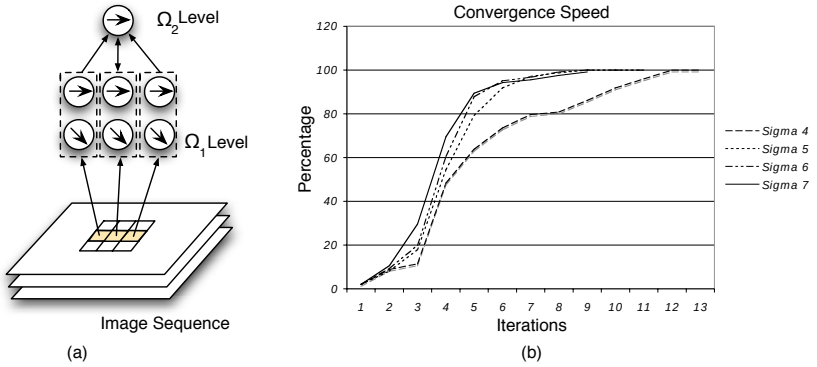

Fig. 2: The two layer structure (a) for $\sigma = 1.5$, connections representing Eq. 2 are not drawed. The convergence of the system to the right detection (b) using different values for σ in Eq. 4. Higher values for σ increase the speed of convergence.

4 Results

The criteria we use to measure our results follow the idea proposed by Barron et al [2]. Fig. 3 shows our testing with a solid untextured square that moves diagonally, with a velocity $\vec{v} = (1,1)$ in pixels by frame. Noise is added close to the moving edge in different orientations, but with magnitudes lower than the true motion, see Fig. 3(a). Our results show how many detectors give the right direction as function of the number of iterations (see Fig. 2(b)).

As initial filtering we use energy based filters with separable Gabor functions (x-y-t) with a single spatio-temporal band [8], in a sequence of 3 images. Experiments are performed by sequentially updating layers Ω_1 and Ω_2, but within each layer updates are randomly made (asynchronous). If synchronous updates are used, an additional stage should be added to Ω_2 as in [9]. Both layers are initialized with the filtering output and an additional noise (see Fig. 3).

ESANN'2008 proceedings, European Symposium on
Artificial Neural Networks - Advances in Computational Intelligence and Learning
Bruges (Belgium), 23-25 April 2008, d-side publi., ISBN 2-930307-08-0.

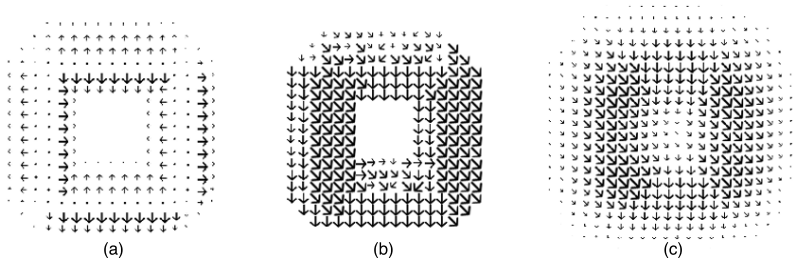

Fig. 3: The filtering response of a diagonally moving square (a) initial filtering with added noise. The layer Ω_1 (b) and Ω_2 (c) after 4 iterations. All three images represent the most activated motion detector orientation for each location.

5 Discussion

Many models inspired by brain circuitry have been proposed for the local detection of motion. All these local filters are sensible to the aperture problem, but also to different kinds of noise: false positive activation, non binary responses (higher responses at right velocities but non-zero ones in other velocities) and contrast variances among others. In our work we propose a model that mainly handles false positive activations within the context of unambiguous motion detection even in very noisy scenario, without losing any spatial precision. Future works include (1) using more realistic noise coming from real image sequences and (2) including static segmentation information in the optical flow extraction.

References

[1] C. Castellanos-Sánchez and B. Girau. Dynamic pursuit with a bio-inspired neural model. In *ACIVS*, pages 284–291, 2005.

[2] J. L. Barron, D. J. Fleet, S. S. Beauchemin, and T. A. Burkitt. Performance of optical flow techniques. *CVPR*, 92:236–242, 1994.

[3] E. P. Simoncelli and D. J. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38(5):743–761, March 1998.

[4] M. A. Giese and T. Poggio. Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, 4(3):179–192, March 2003.

[5] A. Movshon, E. Adelson, M. Gizzi, and W. Newsome. The analysis of moving visual patterns. *Experimental Brain Research*, 11:117–152, 1986.

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

[7] J. P. H. Van Santen and G. Sperling. Elaborated reichardt detectors. *Journal of the Optical Society of America A*, 2(2):300, 1985.

[8] D. J. Heeger. Model for the extraction of image flow. *Journal of the Optical Society of America A*, 4(8):1455–1471, August 1987.

[9] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(10):2041–2066, 2004.