Adaptive Motion Pooling and Diffusion for Optical Flow

N. V. Kartheek Medathati, Manuela Chessa, Guillaume S. Masson, Pierre Kornprobst, Fabio Solari

Adaptive Motion Pooling and Diffusion for Optical Flow

N. V. Kartheek Medathati, Manuela Chessa, Guillaume S. Masson, Pierre Kornprobst, Fabio Solari

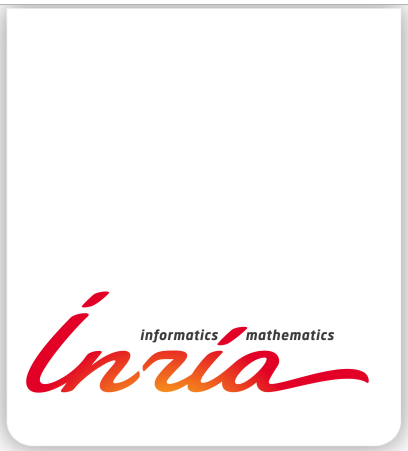
RESEARCH REPORT

N° 8695

March 2015

Project-Team Neuromathcomp

ISSN 0249-6399 ISRN INRIA/RR--8695--FR+ENG



Adaptive Motion Pooling and Diffusion for Optical Flow

N. V. Kartheek Medathati[*†], Manuela Chessa[‡], Guillaume S. Masson[§], Pierre Kornprobst[†], Fabio Solari[§]

Project-Team Neuromathcomp

Research Report n° 8695 — March 2015 — 16 pages

Abstract: We study the impact of local context of an image (contrast and 2D structure) on spatial motion integration by MT neurons. To do so, we revisited the seminal work by Heeger and Simoncelli [40] using spatio-temporal filters to estimate optical flow from V1-MT feedforward interactions (*Feedforward V1-MT model*, FFV1MT). However, the FFV1MT model cannot deal with several problems encountered in real scenes (e.g., blank wall problem and motion discontinuities). Here, we propose to extend the FFV1MT model with adaptive processing by focussing on the role of local context indicative of the local velocity estimates reliability. We set a network structure representative of V1, V2 and MT. We incorporate three functional principles observed in primate visual system: contrast adaptation [38], adaptive afferent pooling [19] and MT diffusion that are adaptive dependent upon the 2D image structure (*Adaptive Motion Pooling and Diffusion*, AMPD). We compared both FFV1MT and AMPD performance on Middlebury optical flow estimation dataset [3]. Our results show that the AMPD model performs better than the FFV1MT model and its overall performance is comparable with many computer vision methods.

Key-words: Bio-inspired approach, optical flow, spatio-temporal filters, motion energy, contrast adaptation, population code, V1, V2, MT, Middlebury dataset

* Both authors N. V. Kartheek Medathati and Manuela Chessa should be considered as first author.
† INRIA, Neuromathcomp team, Sophia Antipolis, France
‡ University of Genova, DIBRIS, Italy
§ Institut de Neurosciences de la Timone, CNRS, Marseille, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Integration et Diffusion Adaptative pour le Calcul du Flot Optique

Résumé : Nous étudions l'impact du contexte local d'une image (contraste et structure 2D) sur l'intégration de mouvement spatial par les neurones MT. Pour ce faire, nous avons revisité le modèle de référence de Heeger et Simoncelli [40] utilisant des filtres spatio-temporels et une intégration entre V1 et MT pour estimer le flot optique (*Feedforward V1-MT model*, FFV1MT). Cependant, le modèle FFV1MT ne permet pas de résoudre plusieurs problèmes rencontrés dans des scènes réelles (par exemple, les problèmes des régions sans contraste et des discontinuités de mouvement). Ici, nous proposons d'étendre le modèle FFV1MT avec traitement adaptatif en mettant l'accent sur le rôle du contexte local indicatif de la fiabilité des estimations de vitesse locale. Nous établissons une structure de réseau de neurones représentant les activités dans les aires corticales V1, V2 et MT. Nous intégrons trois principes fonctionnels observés dans le système visuel des primates: l'adaptation au contraste [38], une intégration adaptative [19] et une diffusion adaptative interne à MT qui dépend de la structure 2D de l'image (*Adaptive Motion Pooling and Diffusion*, AMPD). Nous comparons les deux modèles FFV1MT et AMPD sur la base de test Middlebury [3]. Nos résultats montrent que le modèle AMPD est plus performant que le modèle FFV1MT et que sa performance globale est comparable à de nombreuses méthodes de vision par ordinateur.

Mots-clés : Approche bio-inspirée, flot optique, filtres spatio-temporels, énergie de mouvement, contrat adaptation, codage en population, V1, V2, MT, base de test Middlebury

Contents

1 Introduction

Dense optical flow estimation is a well studied problem in computer vision with several algorithms being proposed and benchmarked over the years [4, 2, 12]. Given that motion information can be used for serving several functional tasks such as navigation, tracking and segmentation, biological systems have evolved sophisticated and highly efficient systems for visual motion information analysis. Understanding the mechanisms adapted by biological systems would be very beneficial for both scientific and technological reasons and has spurred a large number of researchers to investigate underlying neural mechanisms (for reviews see [26, 43, 13, 8, 10]).

Psychophysical and neurophysiological results on global motion integration in primates have inspired many computational models of motion processing. These models first, and still mainly, focus on simulating neural responses to gratings and plaids patterns, are prototypical properties of the local motion estimation stage (area V1) and the motion integration stage (area MT) [25, 46, 40, 35, 44]. However, gratings and plaids are spatially homogeneous motion inputs such that spatial and temporal aspects of motion integration have been largely ignored by these linear-nonlinear filtering models. Dynamical models have been proposed [16, 45, 5] to study these spatial interactions and how they can explain the diffusion of non-ambiguous local motion cues. For instance, [7, 34, 42, 9] have simulated how form-related features in the image such as the luminance gradient or the binocular disparity can play a role in improving motion estimation. However, all these previous models are based on sets of fixed spatiotemporal filters and static nonlinearities. This depart from the highly adaptive and nonlinear properties of visual receptive fields (RFs) [14, 39]. Moreover, these bio-inspired models are barely evaluated in terms of their efficacy on modern computer vision datasets with the notable exceptions such as in [4] (with an early evaluation of spatio-temporal filters) or in [6, 9] (with evaluations on Yosemite or Middlebury videos subset).

In this paper, we propose to fill the gap between studies in biological and computer vision for motion estimation by building our approach on results from visual neuroscience and thoroughly evaluating the method using standard computer vision dataset (Middlebury). The paper is organized as follows. In Sec. 2, we present a brief overview of the motion processing pathway of the primate brain on which our model is based, computational issues to be dealt for the optical flow and known functional principles used by the system to solve them. In Sec. 3, we describe a baseline model for optical flow estimation based on a V1-MT feedforward interactions. This model designated by FFV1MT is largely inspired from Heeger [17, 40] (see [41] for more details). In Sec. 4, the FFV1MT model is extended by taking into account both image structure and contrast adaptive pooling and ambiguity resolution through lateral interactions among MT neurons. In order to handle the scale a typical scale-space pyramidal approach has been considered. In Sec. 5, the proposed model is evaluated using the standard Middlebury dataset. In Sec. 6, we conclude by highlighting our main contributions and we relate our work to standard and recent ideas from computer vision.

2 Biological vision solution

2.1 Cortical hierarchy

In visual neuroscience, properties of low-level motion processing have been extensively investigated in humans [27] and monkeys [28]. Figure 1 illustrates a schematic view of the multiple stages of motion processing in primate cortex. Local motion information is extracted locally through a set of spatiotemporal filters tilling the retinotopic representation of the visual field in area V1. However, these direction-selective cells exhibit several nonlinear properties as the center response is constantly modulated by the surrounding inputs conveyed by feedback and lateral inputs. Direction-selective cells project directly to the motion integration stage. Neurons in the area MT pool these responses over a broad range of spatial and temporal scales, becoming able to extract the direction and speed of a particular surface, regardless

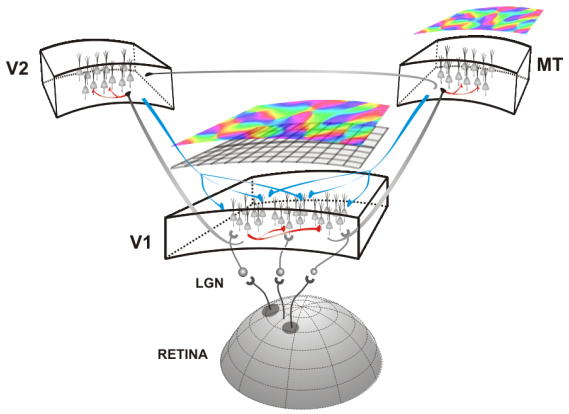


Figure 1: Illustration of the motion processing pathway, with the main cortical areas involved in motion estimation (area V1) and integration (area MT). Interactions with the form pathway are represented by V2 and V4 cortical areas. This cartoon illustrates the variety of connectivities: feedforward (in gray), short- and long-range lateral (in red) and feedback (in blue).

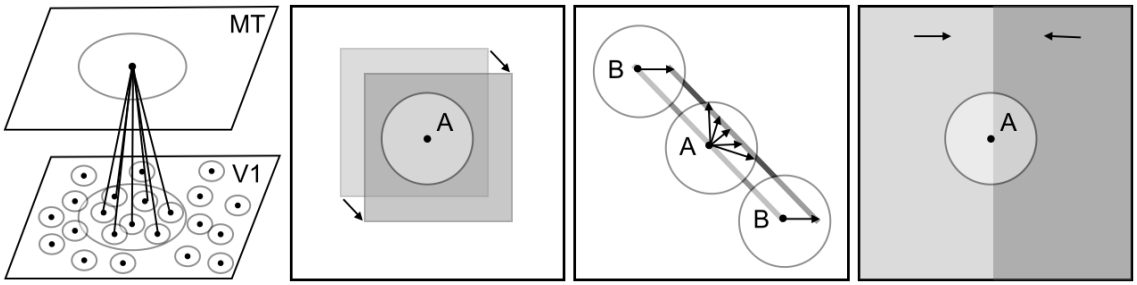
its shape or color [10]. Context modulations are not only implemented by center-surround interactions in areas V1 and MT. For instance, other extra-striate areas such as V2 or V4 project to MT neurons to convey information about the structure of the visual scene, such as the orientation or color of local edges.

2.2 Receptive fields: a local analysis

For each neuron in the hierarchy, one can associate a receptive field defined by the region in the visual field that elicits an answer. Receptive fields are first small and become larger going deeper in the hierarchy [28]. The first local analysis of motion is done at the V1 cortical level. The small receptive field size of V1 neurons, and their strong orientation selectivity, poses several difficulties when estimating global motion direction and speed, as illustrated in Fig. 2. In particular, any local motion analyzer will face the three following computational problems [10]:

- Absence of illumination contrast is referred to as blank wall problem in which the local estimator is oblivious to any kind of motion (Fig. 2(b)).
- Presence of luminance contrast changes along only one orientation is often referred to as aperture problem where the local estimator cannot recover the velocity component along the gradient (Fig. 2(c)).
- Presence of multiple motions or multiple objects with in the receptive field in which case the local estimator has to be selective to arrive at an accurate estimation (Fig. 2(c)).

In terms of optical flow estimation, feedforward computation involving V1 and MT could be sufficient in the case of regions without any ambiguity. On the contrary, recovering velocity at regions where there is some ambiguity such as aperture or blank wall problems imply to pool reliable information from other, less ambiguous regions in the surrounding. Such spatial diffusion of information is thought to be conveyed by the intricate network of lateral (short-range, or recurrent networks, and long-range) (see [20] for reviews).



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

Figure 2: Estimation motion from local observations: (a) Illustration of a pooling step with the corresponding receptive fields; (b) blank wall problem: at position A, the absence of texture gives no information to estimate motion; (c) aperture problem: at position A, only the 1D component of the flow is known; (d) multiple motion: at position A, receptive field integrates different motion informations. Problems in motion estimation.

2.3 Contrast adaptive processing

The structure of neuronal receptive fields is not static as it has long been thought [14]. Rather, it adapts to the local context of the image so that many of the tuning functions characterizing low-level neurons are in fact dynamical [39]. A first series of evidence comes from experiments where the properties of the local inputs change the classical receptive field. For instance orientation-tuning in area V1 and speed tuning of MT neurons are sharper when tested with broad-band texture inputs, as compared to low-dimension gratings [15, 32]. Moreover, spatial summation function often broadens as contrast decreases or noise level increases [38]. These observations are complemented by experiments varying the spatial context of this local input. For instance, surround inhibition in V1 and MT neurons becomes stronger at high contrast and center-surround interactions exhibit a large diversity in terms of their relative tunings. Moreover, the spatial structure of these interactions is often more diverse in shape that the classical view of the concentric, Mexican-hat areas (see [10] for a review). Lastly, at each decoding stage, it seems nowadays that tuning functions are weighted by the reliability of the neuronal responses, as varying for instance with contrast or noise levels [23]. Still, these highly adaptive properties have barely been taken into account when modeling visual motion processing. Here, we model some of these mechanisms to highlight their potential impact on optic flow computation. We focus on both the role of local image structure (contrast, textureness) and the reliability of these local measurements in controlling the diffusion mechanisms. We investigated how these mechanisms can help solving local ambiguities, and segmenting the flow fields into different surfaces while still preserving the sharpness and precision of natural vision.

3 Baseline Model (FFV1MT)

In this section we briefly introduce the FFV1MT model introduced in [41], in which we revisited the seminal work by Heeger [17, 40] using spatio-temporal filters to estimate optical flow. FFV1MT model is a three-step approach, corresponding to area V1, area MT and decoding of MT response.

In term of notations, we consider a grayscale image sequence $I(x, y, t)$, for all positions $p = (x, y)$ inside a domain Ω and for all time $t > 0$. Our goal is to find the optical flow $v(x, y, t) = (v_x, v_y)(x, y, t)$ defined as the apparent motion at each position p and time t.

3.1 Area V1: Motion Energy

Area V1 comprises simple and complex cells to estimate motion energy [1]. Complex cells receive inputs from several simple cells and their response properties have been modelled by the motion energy, which is a non linear combination of afferent simple cell responses.

Simple cells are characterized by the preferred direction θ of their contrast sensitivity in the spatial domain and their preferred velocity v^c in the direction orthogonal to their contrast orientation often referred to as component speed. The RFs of the V1 simple cells are classically modeled using band-pass filters in the spatio-temporal domain. In order to achieve low computational complexity, the spatio-temporal filters are decomposed into separable filters in space and time. Spatial component of the filter is described by Gabor filters h and temporal component by an exponential decay function k. Given a spatial size of the receptive field σ and the peak spatial and temporal frequencies f_s and f_t, we define the following complex filters by:

$$h(p,\theta,f_s) = Be^{\left(\frac{-(x^2+y^2)}{2\sigma^2}\right)} e^{j2\pi(f_s cos(\theta)x + f_s sin(\theta)y)}, \tag{1}$$

$$k(t,f_t) = e^{\left(-\frac{t}{\tau}\right)} e^{j2\pi(f_t t)}, \tag{2}$$

where σ and τ are the spatial and temporal scales respectively. Denoting the real and imaginary components of the complex filters h and p as h_e, k_e and h_o, k_o respectively, and a preferred velocity v_c related to the frequencies by the relation

$$v^c = \frac{f_t}{f_s}, \tag{3}$$

we introduce the odd and even spatio-temporal filters defined as follows,

$$\begin{aligned} g_o(p,t,\theta,v^c) &= h_o(p,\theta,f_s)k_e(t;f_t) + h_e(p,\theta,f_s)k_o(t;f_t), \\ g_e(p,t,\theta,v^c) &= h_e(p,\theta,f_s)k_e(t;f_t) - h_o(p,\theta,f_s)k_o(t;f_t). \end{aligned}$$

These odd and even symmetric and tilted (in space-time domain) filters characterize V1 simple cells. Using these expressions, we define the response of simple cells, either odd or even, with a preferred direction of contrast sensitivity θ in the spatial domain, with a preferred velocity v^c and with a spatial scale σ by

$$R_{o/e}(p,t,\theta,v^c) = g_{o/e}(p,t,\theta,v^c) \overset{(x,y,t)}{*} I(x,y,t) \tag{4}$$

The complex cells are described as a combination of the quadrature pair of simple cells (4) by using the motion energy formulation,

$$E(p,t,\theta,v^c) = R_o(p,t,\theta,v^c)^2 + R_e(p,t,\theta,v^c)^2,$$

followed by a normalization. Assuming that we consider a finite set of orientations $\theta = \theta_1 \dots \theta_N$, to obtain the final V1 response

$$E^{V1}(p,t,\theta,v^c) = \frac{E(p,t,\theta,v^c)}{\sum_{i=1}^{N} E(p,t,\theta_i,v^c) + \varepsilon}, \tag{5}$$

where $0 < \varepsilon \ll 1$ is a small constant to avoid divisions by zero in regions with no energies which happen when no spatio-temporal texture is present. The main property of V1 is its tuning to the spatial orientation of the visual stimulus,since the preferred velocity of each cell is related to the direction orthogonal to its spatial orientation.

3.2 Area MT: Pattern Cells Response

MT neurons exhibit velocity tuning irrespective of the contrast orientation. This is believed to be achieved by pooling afferent responses in both spatial and orientation domains followed by a non-linearity. Our modelling of area MT comprises pattern cells [40, 36]. The responses of an MT pattern cell tuned to the speed v^c and to direction of speed d can be expressed as follows:

$$E^{MT}(p,t;d,v^c) = F\left(\sum_{i=1}^{N} w_d(\theta_i)\mathscr{P}(E^{V1})(p,t;\theta_i,v^c)\right),$$

where w_d represents the MT linear weights that give origin to the MT tuning and $\mathscr{P}(E^{V1})$ corresponds to the spatial pooling.

The physiological evidence suggests that w_d is a smooth function with central excitation and lateral inhibition. Cosine function shifted over various orientations is a potential function that could satisfy this requirement to produce the responses for a population of MT neurons [24]. Considering the MT linear weights shown in [36], $w_d(\theta)$ is defined by θ:

$$w_d(\theta) = \cos(d-\theta) \quad d \in [0, 2\pi[. \tag{6}$$

This choice allows us to obtain direction tuning curves of pattern cells that behave as in [36].

The spatial pooling term is defined by

$$\mathscr{P}(E^{V1})(p,t;\theta_i,v^c) = \frac{1}{N}\sum_{p'} f_\alpha(\|p-p'\|)E^{V1}(p,t;\theta_i,v^c) \tag{7}$$

where $f_\mu(s) = \exp(s^2/2\mu^2)$, $\|.\|$ is the L_2-norm, α is a constant, $\bar{N}$ is a normalization term (here equal to $2\pi\alpha^2$) and $F(s) = exp(s)$ is a static nonlinearity chosen as an exponential function [29, 36]. The pooling defined by (7) is a simple spatial Gaussian pooling.

3.3 Sampling and Decoding MT Response: Optical Flow Estimation

In order to engineer an algorithm capable of recovering dense optical flow estimates, we still need to address problems of sampling and decoding the population responses of heterogeneously tuned MT neurons. In [41], we proposed a new decoding stage to obtain a dense optical flow estimation from the MT population response. Note that, although this model is mostly classical, it had never been tested before on modern computer vision datasets.

Indeed, a unique velocity vector cannot be recovered from the activity of a single velocity tuned MT neuron as multiple scenarios could evoke the same activity. However, a unique vector can be recovered from the population activity. In this section, we present a decoding step which was not present in [40, 36] to decode the MT population.

The velocity space could be sampled by considering MT neurons that span over the 2-D velocity space with a preferred set of tuning speed directions in $[0, 2\pi[$ and also a multiplicity of tuning speeds. Sampling the whole velocity space is not required, as a careful sampling along the cardinal axes could be sufficient to recover the full velocity vector.

In this paper, we sample the velocity space using two MT populations tuned to the directions $d = 0$ and $d = \pi/2$ with varying tuning speeds. Here, we adopt a simple weighted sum approach to decode the MT population response [33].

$$\begin{cases} v_x(p,t) = \sum_{i=1}^{M} v_i^c E^{MT}(p,t,0,v_i^c), \\ v_y(p,t) = \sum_{i=1}^{M} v_i^c E^{MT}(p,t,\pi/2,v_i^c). \end{cases} \tag{8}$$

Note that other decoding methods exist such as, e.g., the maximum likelihood estimator [31, 30], however we have adopted the linear weighted sum approach, as a balancing choice between simplicity, computational cost and reliability of the estimates.

4 Adaptive Motion Pooling and Diffusion Model (AMPD)

The baseline model FFV1MT involving a feedforward processing from V1 to MT is largely devised to describe physiological and psychophysical observations on motion estimation when the testing stimuli were largely homogeneously textured regions such as moving gratings and plaids. Hence the model is limited in the context of dense flow estimation for natural videos as it has no inherent mechanism to deal with associated sub problems such blank wall problem, aperture problem or occlusion boundaries.

Building on recent results summarized in Sec. 2.3 we model some of these mechanisms to highlight their potential impact on optic flow computation. Considering inputs from area V2, we focus on the role of local context (contrast and image structure) indicative of the reliability of these local measurements in (i) controlling the pooling from V1 to MT and (ii) adding lateral connectivity in MT.

4.1 Area V2: Contrast and Image Structure

Our goal is to define a measure of contrast which is indicative of the aperture and blank wall problems using the responses of spatial Gabor filters. There exist several approaches to characterize the spatial content of an image from Gabor filter. For example, in [21] the authors propose the phase congruency approach which detects edges and corners irrespectively of contrast in an image. In dense optical flow estimation problem, region with texture are less likely to suffer blank wall and aperture problems even though edges are susceptible to aperture problem. So phase congruency approach cannot be used directly and we propose the following simple alternative approach.

Let h_{θ_i} the Gabor filter for edge orientation θ_i, we define

$$R(p) = (R_{\theta_1}(p), \ldots, R_{\theta_N}(p)) \text{ where } R_{\theta_i}(p) = |h_{\theta_i} * I|(p).$$

Given an edge orientation at θ_i, R_{θ_i} is maximal when crossing the edge and ∇R_{θ_i} indicate the direction to go away from edge.

Then the following contrast/cornerness measure is proposed as follows, taking into consideration the amount of contrast at a given location and also ensuring that contrast is not limited to a single orientation giving raise to aperture problem.

$$\mu(R)(p) = \frac{1}{N} \sum_i R_{\theta_i}(p), \tag{9}$$

$$C(p) = H_\xi(\mu(R(p))(1 - \sigma^2(R(p))/\sigma^2_{max}), \tag{10}$$

where $\mu(R(p))$ (resp. $\sigma^2(R(p))$) denote the average (resp. variance) of components of R at position p, $H_\xi(s)$ is a step function ($H_\xi(s) = 0$ if $s \leq \xi$ and 1 otherwise) and $\sigma^2_{max} = \max_{p'} \sigma^2(R(p))$. The term $H_\xi(\mu(R(p))$ is an indicator of contrast as it measures the Gabor energies: in regions with strong contrast or strong texture in any orientation this term equals to one; in a blank wall situation, it is equal to zero. The term $(1 - \sigma^2(R(p))/\sigma^2_{max})$ measures how strongly the contrast is oriented in a single direction: it is higher when there is only contrast in one direction and lower when there is contrast in more than one orientation (thus it is an indicator of where there is aperture problem).

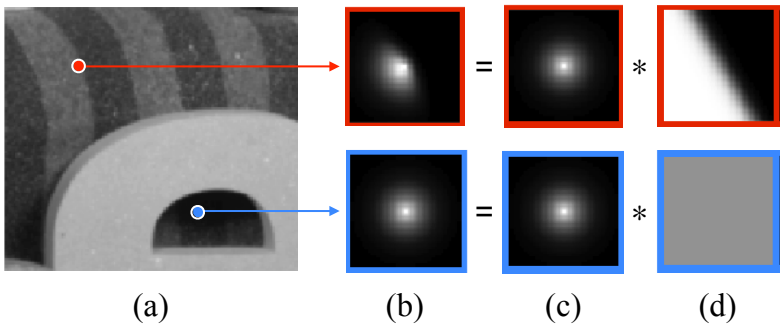


Figure 3: Example of pooling weights at different positions: (a) Sample input indicating two different positions p (see red and blue dots) at which we show: (b) the final pooling weight $W(\cdot, p)$ which is obtained by multiplying (c) the isotropic term by the (d) anisotropic term (see text).

4.2 Area MT: V2-Modulated Pooling

Most of the models currently pool V1-afferents using a linear fixed receptive field size, which does not adapt itself to the local gradient or respect discontinuities in spatio-temporal reposes. This might lead to degradation in the velocity estimates by blurring edges/kinetic boundaries. Thus it is advantageous to make the V1 to MT pooling adaptive as a function of texture edges.

We propose to modify the pooling stage as follows

$$E^{MT}(p,t;d,v^c) = F\left(\sum_{i=1}^{N} w_d(\theta_i)\tilde{\mathscr{P}}(E^{V1})(p,t;\theta_i,v^c)\right),$$

where the spatial pooling become functions of image structure. We propose the following texture-dependent spatial pooling:

$$\mathscr{P}(E^{V1})(p,t;\theta_i,v^c) = \frac{1}{\bar{N}(p,\theta_i)}\sum_{p'} W(p,p')E^{V1}(p,t;\theta_i,v^c), \tag{11}$$

$$\text{where } W(p,p') = f_{\alpha(\|R\|(p))}(\|p-p'\|)g_i(p,p'),$$

and where $\bar{N}(p,\theta_i) = \sum_{p'} W(p,p')$ is a normalizing term. Note that the weight $W(p,p')$ has two components which depend on image structure as follows. Term $f_{\alpha(\|R\|(p))}(\|p-p'\|)$ is an isotropic weight setting the size of the integration domain. The variance of the distance term α depends on the structure R_{θ_i}:

$$\alpha(\|R\|(p)) = \alpha_{max}e^{-\eta\frac{\|R\|^2(p)}{r_{max}}}, \tag{12}$$

where η is a constant, $r_{max} = \max_{p'}\{\|R\|^2(p')\}$. Term $g_i(p,p')$ is an anisotropic weight enabling anisotropic pooling close to image structures so that discontinuities could be better preserved. Here we propose to define g_i by

$$g_i(p,p') = S_{\lambda,\nu}\left(-\frac{\nabla R_{\theta_i}(p)}{\|\nabla R_{\theta_i}\|+\varepsilon}\cdot(p'-p)\right), \tag{13}$$

where $S_{\lambda,\nu} = 1/(1+\exp(-\lambda(x-\nu)))$ is a sigmoid function and ε a small constant. Note that this term is used only in regions where $\|\nabla R_{\theta_i}\|$ is greater than a threshold. Fig. 3 gives two examples of the pooling coefficients at different positions.

4.3 MT Lateral Interactions

We model the lateral iterations for the velocity information spread (from the regions where there is less ambiguity to regions with high ambiguity, see Sec. 2.2) whilst preserving discontinuities in motion and illumination. To do so, we propose an iterated trilateral filtering defined by:

$$u^{n+1}(p) = \frac{1}{\bar{N}(p)} \sum_{p'} W(p,p')u^n(p'), \tag{14}$$

$$c^{n+1}(p) = c^n(p) + \lambda(\max_{p' \in \mathcal{N}(p)} c^n(p') - c^n(p)) \tag{15}$$

$$u^0(p) = E^{MT}(p,t;\theta_i,v^c), \tag{16}$$

$$c^0(p) = C(p), \tag{17}$$

where

$$W(p,p') = c^n(p')f_\alpha(\|p-p'\|)f_\beta(c^n(p)(u^n(p')-u^n(p)))$$
$$f_\gamma(I(p')-I(p))u^n(p'), \tag{18}$$

and $\mathcal{N}(p)$ is a local neighborhood around p. The term $c(p')$ ensures that more weight is given naturally to high confidence estimates; The term $c(p)$ inside f_β ensures that differences in the MT responses are ignored when confidence is low facilitating the diffusion of information from regions with high confidence and at the same time preserves motion discontinuities or blurring at the regions with high confidence.

5 Results

In order to test the method a multi-scale version of both the baseline approach FFV1MT and approach with adaptive pooling AMPD are considered. The method is applied on a Gaussian pyramid with 6 scales, the maximum number of scales that could be reliably used for the spatio-temporal filter support that has been chosen.

A first test was done on Yosemite sequence (without clouds) as it is widely used in both computer vision and biological vision studies (see Fig. 4, first row). For FFV1MT, AEE=3.55 $\pm$ 2.92 and for AMPD we have AAE=3.00 $\pm$ 2.21. This can be compared to what has been obtained with previous biologically-inspired models such as the original Heeger approach (AAE=11.74°, but estimated 44.8% of the most reliable regions, see [4]) and the neural model from Bayerl and Neumann (AAE=6.20°, [5]), showing an improvement. One can do comparisons with standard computer vision approaches such as Pyramidal Lucas and Kanade (AAE=6.41°) and Horn and Schunk (AAE=4.01°, [18]), showing a better performance.

The results on the Middlebury training set show improvements of the proposed method with respect to FFV1MT (see Table 1). For qualitative comparison, sample results are also presented in Fig.4. The relative performance of extended method can be understood by observing δAAE, difference between the FFV1MT AAE map and the AMPD AAE map which are presented in Fig. 4 (last column): the improvements are prominent at the edges, e.g., see the δAAE column for the RubberWhale and Urban2 sequence.

6 Conclusion

In this paper, we have proposed a new algorithm that incorporates three functional principles observed in primate visual system, namely contrast adaptation, image structure based afferent pooling and ambiguity

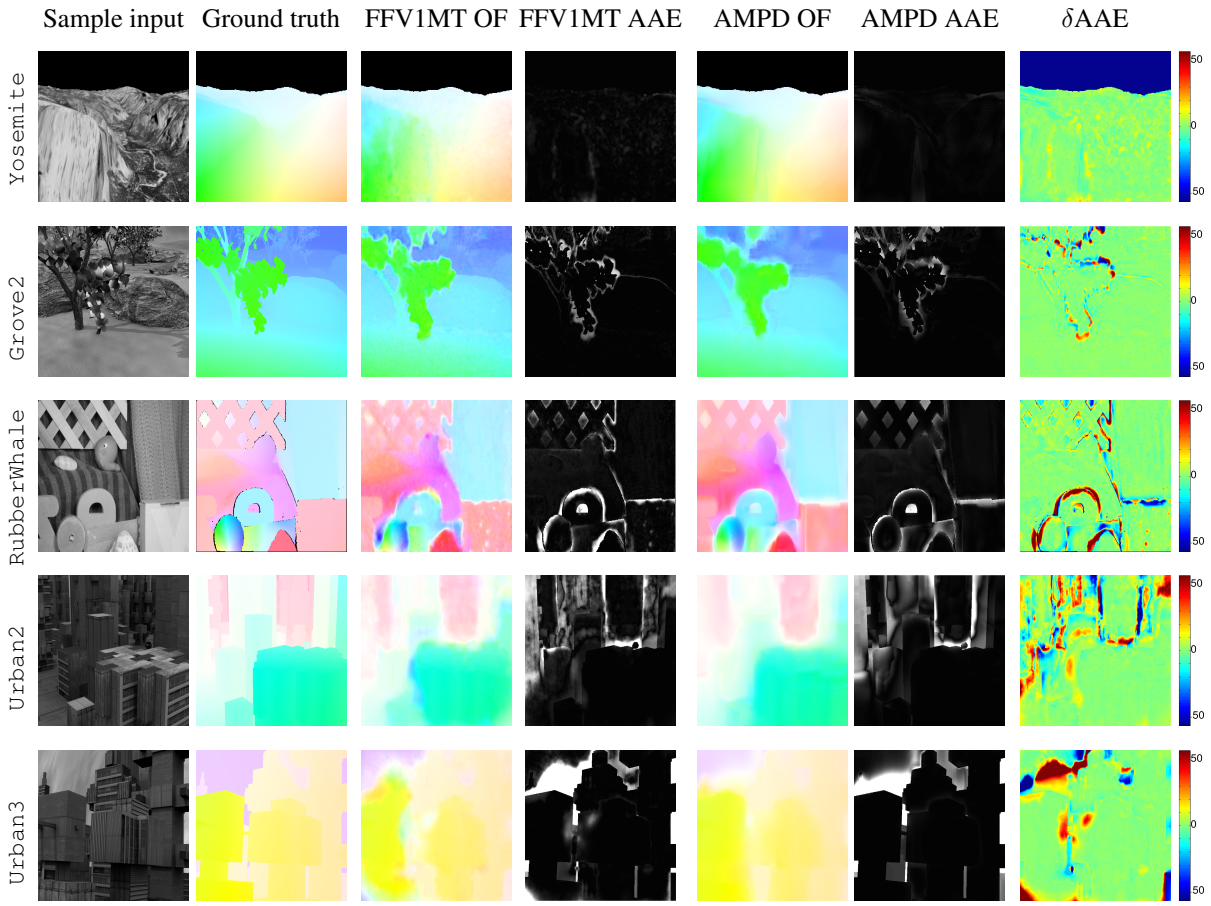


Figure 4: Sample results on Yosemite sequence and a subset of Middlebury training set. δAAE = AAE_{FFV1MT} - AAE_{AMPD}

	FFV1MT		AMPD	
Sequence	AAE $\pm$ STD	EPE $\pm$ STD	AAE $\pm$ STD	EPE $\pm$ STD
grove2	4.28 $\pm$ 10.25	0.29 $\pm$ 0.62	4.07 $\pm$ 9.29	0.27 $\pm$ 0.56
grove3	9.72 $\pm$ 19.34	1.13 $\pm$ 1.85	10.66 $\pm$ 19.25	1.11 $\pm$ 1.61
Hydrangea	5.96 $\pm$ 11.17	0.62 $\pm$ 0.96	5.48 $\pm$ 11.10	0.50 $\pm$ 0.69
RubberWhale	10.20 $\pm$ 17.67	0.34 $\pm$ 0.54	8.87 $\pm$ 13.16	0.30 $\pm$ 0.42
urban2	14.51 $\pm$ 21.02	1.46 $\pm$ 2.13	12.70 $\pm$ 19.92	1.09 $\pm$ 1.31
urban3	15.11 $\pm$ 35.28	1.88 $\pm$ 3.27	12.78 $\pm$ 31.36	1.32 $\pm$ 2.25

Table 1: Error measurements on Middlebury training set

based lateral interaction. This is an extension to an earlier algorithm FFV1MT [41] inspired by Heeger et al. [17, 40] which is appreciated by both computer vision and biological vision communities.

Contemporary computer vision methods to [17] such as [22] and [18] which study local motion estimation and global constraints to solve aperture problem have been revisited by the computer vision with great interest [11] and a lot of investigations are being carried out to regulate the information diffusion

from non-ambiguous regions to ambiguous regions based on image structure (see, e.g., [37]). Very few attempts have been made to incorporate these ideas into spatio-temporal filter based models such as [17]. Given the recent growth in neuroscience, it is very interesting to revisit this model incorporating the new findings and examining the efficacy.

This paper provides a baseline for future research in biologically-inspired computer vision, considering that functional principles uncovered in biological vision are a rich source of ideas for future developments. The extended model AMPD improved the flow estimation compared to FFV1MT and has a great potential to be further improved. It has opened up several interesting sub problems, which could be of relevance to biologists as well, for example to investigate what could be afferent pooling strategy of MT when there are multiple surfaces or occlusion boundaries within the MT receptive field, or if we could recover a better dense optical flow map by considering decoding problem as a deblurring problem due the spatial support of the filters.

Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318723 (MATHEMACS) and grant agreement no. 269921 (BrainScaleS).

References

[1] E. Adelson and J. Bergen. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2:284–299, 1985. 7

[2] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *International Conference on Computer Vision, ICCV'07*, pages 1–8, 2007. 4

[3] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011. 1, 2

[4] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. *The International Journal of Computer Vision*, 12(1):43–77, 1994. 4, 11

[5] P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16(10):2041–2066, 2004. 4, 11

[6] P. Bayerl and H. Neumann. A fast biologically inspired algorithm for recurrent motion estimation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(2):246–260, 2007. 4

[7] C. Beck and H. Neumann. Interactions of motion and form in visual cortex – a neural model. *Journal of Physiology - Paris*, 104:61–70, 2010. 4

[8] R. Born and D. Bradley. Structure and function of visual area MT. *Annu. Rev. Neurosci*, 28:157–189, 2005. 4

[9] J. Bouecke, E. Tlapale, P. Kornprobst, and H. Neumann. Neural mechanisms of motion detection, integration, and segregation: From biology to artificial image processing systems. *EURASIP Journal on Advances in Signal Processing*, 2011, 2011. special issue on Biologically inspired signal processing: Analysis, algorithms, and applications. 4

[10] D. Bradley and M. Goyal. Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 9(9):686–695, 2008. 4, 5, 6

[11] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61:211–231, 2005. 12

[12] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part VI*, ECCV'12, pages 611–625, Berlin, Heidelberg, 2012. Springer-Verlag. 4

[13] C. Clifford and M. Ibbotson. Fundamental mechanisms of visual motion detection: models, cells and functions. *Progress in Neurobiology*, 68(6):409 – 437, 2002. 4

[14] A. Fairhall. The receptive field is dead. long life the receptive field? *Current Opinion in Neurobiology*, 25:9–12, 2014. 4, 6

[15] J. Freeman, C. M. Ziemba, D. J. Heeger, E. P. Simoncelli, and J. A. Movshon. A functional and perceptual signature of the second visual area in primates. *Nature Neuroscience*, 16:974–981, 2013. 6

[16] S. Grossberg and E. Mingolla. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychological review*, 92(2):173–211, 1985. 4

[17] D. Heeger. Optical flow using spatiotemporal filters. *The International Journal of Computer Vision*, 1(4):279–302, Jan. 1988. 4, 6, 12, 13

[18] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981. 11, 12

[19] X. Huang, T. Albright, and G. Stoner. Adaptive surround modulation in cortical area MT. *Neuron*, 53:761–770, 2007. 1, 2

[20] U. Ilg and G. Masson. *Dynamics of Visual Motion Processing: Neuronal, Behavioral, and Computational Approaches*. SpringerLink: Springer e-Books. Springer Verlag, 2010. 5

[21] P. Kovesi. Image features from phase congruency. *Videre: A Journal of Computer Vision Research. MIT Press*, 1(3), 1999. 9

[22] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981. 12

[23] W. J. Ma and M. Jazayeri. Neural coding of uncertainty and probability. *Annual Review of Neuroscience*, 37(1):205–220, 2014. 6

[24] J. H. Maunsell and D. C. Van Essen. Functional properties of neurons in middle temporal visual area of the macaque monkey. I. selectivity for stimulus direction, speed, and orientation. *Journal of Neurophysiology*, 49(5):1127–1147, 1983. 8

[25] J. Movshon, E. Adelson, M. Gizzi, and W. Newsome. The analysis of visual moving patterns. *Pattern recognition mechanisms*, pages 117–151, 1985. 4

[26] K. Nakayama. Biological image motion processing: A review. *Vision Research*, 25:625–660, 1984. 4

[27] S. Nishida. Advancement of motion psychophysics: Review 2001-2010. *Journal of Vision*, 11(5):11, 1–53, 2011. 4

[28] G. A. Orban. Higher order visual processing in macaque extrastriate cortex. *Physiological Reviews*, 88(1):59–89, 2008. 4, 5

[29] L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network: Computation in Neural Systems*, 15(4):243–262, 2004. 8

[30] A. Pouget, P. Dayan, and R. Zemel. Information processing with population codes. *Nature Reviews Neuroscience*, 1(2):125–132, 2000. 9

[31] A. Pouget, K. Zhang, S. Deneve, and P. E. Latham. Statistically efficient estimation using population coding. *Neural Computation*, 10(2):373–401, 1998. 9

[32] N. Priebe, C. Cassanello, and S. Lisberger. The neural representation of speed in macaque area MT/V5. *Journal of Neuroscience*, 23(13):5650–5661, July 2003. 6

[33] K. R. Rad and L. Paninski. Information rates and optimal decoding in large neural populations. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, editors, *NIPS*, pages 846–854, 2011. 8

[34] F. Raudies and H. Neumann. A model of neural mechanisms in monocular transparent motion perception. *Journal of Physiology-Paris*, 104(1-2):71–83, 2010. 4

[35] N. Rust, V. Mante, E. Simoncelli, and J. Movshon. How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9:1421–1431, 2006. 4

[36] N. C. Rust, V. Mante, E. P. Simoncelli, and J. A. Movshon. How MT cells analyze the motion of visual patterns. *Nature Neuroscience*, 9(11):1421–1431, 2006. 8

[37] J. Sánchez, A. Salgado, and N. Monzón. Preserving accurate motion contours with reliable parameter selection. In *ICIP*, pages 209–213, 2014. 13

[38] M. P. Sceniak, D. L. Ringach, M. J. Hawken, and R. Shapley. Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience*, 2(8):733—739, Aug. 1999. 1, 2, 6

[39] T. Sharpee, H. Sugihara, A. Kurgansky, S. Rebrik, M. Stryker, and K. Miller. Adaptive filtering enhances information transmission in visual cortex. *Nature*, 439:936–942, 2006. 4, 6

[40] E. Simoncelli and D. Heeger. A model of neuronal responses in visual area MT. *Vision Research*, 38:743–761, 1998. 1, 2, 4, 6, 8, 12

[41] F. Solari, M. Chessa, K. Medathati, and P. Kornprobst. What can we expect from a classical v1-mt feedforward architecture for optical flow estimation? *Signal Processing: Image Communication*, 2015. 4, 6, 8, 12

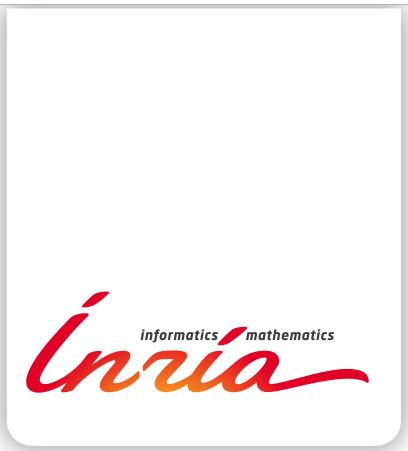
[42] E. Tlapale, G. S. Masson, and P. Kornprobst. Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. *Vision Research*, 50(17):1676–1692, Aug. 2010. 4

[43] D. C. Van Essen and J. L. Gallant. Neural mechanisms of form and motion processing in the primate visual system. *Neuron*, 13:1–10, July 1994. 4

[44] Y. Weiss and E. Adelson. Adventures with gelatinous ellipses – constraints on models of human motion analysis. *Perception*, 29:543–566, 2000. 4

[45] Y. Weiss and E. H. Adelson. Slow and smooth: A Bayesian theory for the combination of local motion signals in human vision. *Center for Biological and Computational Learning Paper*, 1998. 4

[46] H. Wilson, V. Ferrera, and C. Yo. A psychophysically motivated model for two-dimensional motion perception. *Visual Neuroscience*, 9(1):79–97, July 1992. 4



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399