Adaptive Motion Pooling and Diffusion for Optical Flow Computation

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

Adaptive Motion Pooling and Diffusion for Optical Flow Computation

N. V. Kartheek Medathati[1], Manuela Chessa[2], Guillaume S. Masson[3], Pierre Kornprobst[1], and Fabio Solari[2]

[1] INRIA, Biovision team, Sophia Antipolis, France
[2] Department of Informatics, Bioengineering, Robotics and System Engineering - DIBRIS, University of Genoa, Genova, Italy
[3] Institut de Neurosciences de la Timone, CNRS, Marseille, France
{kartheek.medathati,pierre.kornprobst}@inria.fr
{manuela.chessa,fabio.solari}@unige.it
guillaume.masson@univ-amu.fr

Abstract. We propose to extend a state of the art bio-inspired model for optic flow computation through adaptive processing by focusing on the role of local context indicative of the local velocity estimates reliability. We set a network structure representative of cortical areas V1, V2 and MT, and incorporate three functional principles observed in primate visual system: contrast adaptation, adaptive afferent pooling and MT diffusion that are adaptive dependent upon the 2D image structure (*Adaptive Motion Pooling and Diffusion*, AMPD). We assess the AMPD performance on Middlebury optical flow estimation dataset, showing that the proposed AMPD model performs better than the baseline one and its overall performance is comparable with many computer vision methods.

Keywords: Brain-inspired computer vision, optic flow, spatio-temporal filters, motion energy, contrast adaptation, population code, V1, V2, MT, Middlebury dataset

1 Introduction

Dense optical flow estimation is a well studied problem in computer vision with several algorithms being proposed and benchmarked over the years [1, 7]. 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 adopted 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 [5].

Psychophysical and neurophysiological results on global motion integration in primates have inspired many computational models of motion processing[19, 17]. 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 [3] to study these spatial interactions and how they can explain the diffusion of non-ambiguous local motion cues [4]. Moreover, the bio-inspired models [12] are barely evaluated in terms of their efficacy on modern computer vision datasets with the notable exceptions such as in [2] (with an early evaluation of spatio-temporal filters) or in [4] (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). It is worth noting that the main interest of this work is not to compete with the state of the art (resulting from more than 20 years of intense research by computer vision community) but to show where a classical model from neuroscience stands with respect to computer vision approaches. 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, and we describe a state of the art model (i.e. a baseline to be improved) for optical flow estimation based on V1-MT feedforward interactions (see [20] for more details). In Sec. 3, we propose the AMPD model, which extends the baseline one through principles inspired by functions of the visual system of the brain by taking into account both image structure and contrast adaptive pooling and ambiguity resolution through lateral interactions among MT neurons. In Sec. 4, the proposed model is evaluated using the standard Middlebury dataset, and Sec. 5 is left for the conclusion.

2 Biological vision solutions and a state of the art model

Cortical hierarchy In visual neuroscience, properties of low-level motion processing have been extensively investigated in humans and monkeys [13]. Local motion information is extracted locally through a set of spatiotemporal filters in area V1. 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 its shape or color [5]. Context modulations are not only implemented by center-surround interactions in areas V1 and MT, but 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 [12].

Receptive fields: a local analysis Receptive fields (RFs) in the visual field are first small and become larger going deeper in the hierarchy [13]. The small RF size of V1 neurons, and their strong orientation selectivity, poses several difficulties when estimating global motion direction and speed. In particular, any local motion analyzer will face the three following computational problems [5]:

- Absence of illumination contrast is referred to as blank wall problem, in which the local estimator is oblivious to any kind of motion.
- 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.
- Presence of multiple motions or multiple objects with in the RF, in which case the local estimator has to be selective to arrive at an accurate estimation.

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 connections – short-range, or recurrent networks, and long-range – (see [9] for reviews).

Contrast adaptive processing The structure of neuronal RFs adapts to the local context of the image [18], and, 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 [8]. Moreover, spatial summation function often broadens as contrast decreases or noise level increases. 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 different from the Mexican-hat structure [5]. 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. 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.

2.1 Baseline Model (FFV1MT)

In this section, we briefly introduce the FFV1MT model proposed in [20], in which we revisited the seminal work by Heeger [19] 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.

- *Area V1: Motion Energy.* Area V1 comprises simple and complex cells to estimate motion energy. Complex cells receive inputs from several simple cells and their response properties have been modeled 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 modeled using band-pass filters in the spatio-temporal domain: the spatial component of the filter is described by Gabor filters h and temporal component by an exponential decay function k. Denoting the real and imaginary components of the complex filters h and k as h_e, k_e and h_o, k_o respectively, and a preferred velocity v^c we introduce the odd $g_o(p,t,\theta,v^c) = h_o(p,\theta,f_s)k_e(t;f_t) + h_e(p,\theta,f_s)k_o(t;f_t)$, and even $g_e(p,t,\theta,v^c) = h_e(p,\theta,f_s)k_e(t;f_t) - h_o(p,\theta,f_s)k_o(t;f_t)$ spatio-temporal filters, where f_s and f_t denote the peak spatial and temporal frequencies. 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

$$R_{o/e}(p,t,\theta,v^c) = g_{o/e}(p,t,\theta,v^c) \overset{(x,y,t)}{*} I(x,y,t) \tag{1}$$

The complex cells are described as a combination of the quadrature pair of simple cells (1) 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

$$E^{V1}(p,t,\theta,v^c) = \frac{E(p,t,\theta,v^c)}{\sum_{i=1}^{N} E(p,t,\theta_i,v^c) + \varepsilon}, \tag{2}$$

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.
- *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. The responses of an MT pattern cell [19, 17] 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 $w_d(\theta) = \cos(d-\theta),\ d \in [0, 2\pi[$, represents the MT linear weights that give origin to the MT tuning, $F(s) = exp(s)$ is a static nonlinearity chosen as an exponential function [14, 17], and $\mathscr{P}(E^{V1})$ corresponds to the spatial pooling.
Cosine function shifted over various orientations is a potential function that could satisfy this requirement (i.e. smooth function with central excitation

and lateral inhibition) to produce the responses for a population of MT neurons [11]. The spatial pooling term is defined by

$$\mathscr{P}(E^{V1})(p,t;\theta_i,v^c) = \frac{1}{\bar{N}} \sum_{p'} f_\alpha(\|p-p'\|) E^{V1}(p',t;\theta_i,v^c) \tag{3}$$

where $f_\mu(s) = \exp(s^2/2\mu^2)$, $\|.\|$ is the L_2-norm, α is a constant, and $\bar{N}$ is a normalization term (here equal to $2\pi\alpha^2$). The pooling defined by (3) is a simple spatial Gaussian pooling.

- *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 [20], we proposed a new decoding stage to obtain a dense optical flow estimation from the MT population response. 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 [15].

$$\begin{cases} v_x(p,t) = \sum_{i=1}^{M} v_i^c E^{MT}(p,t,0,v_i^c), \\ v_y(p,t) = \sum_{i=1}^{M} v_i^c E^{MT}(p,t,\pi/2,v_i^c). \end{cases} \tag{4}$$

3 Adaptive Motion Pooling and Diffusion Model (AMPD)

The baseline model FFV1MT 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 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.

3.1 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, by using the responses of spatial Gabor filters. There exist several approaches to characterize the spatial content of an image from Gabor filter (e.g., in [10] 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:

$$\mu(R(p)) = \frac{1}{N} \sum_i R_{\theta_i}(p), \tag{5}$$

$$C(p) = H_\xi(\mu(R(p))(1 - \sigma^2(R(p))/\sigma^2_{max}), \tag{6}$$

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).

3.2 Area MT: V2-Modulated Pooling

Most of the models currently pool V1-afferents using a linear fixed RF 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:

$$\tilde{\mathscr{P}}(E^{V1})(p, t; \theta_i, v^c) = \frac{1}{\bar{N}(p, \theta_i)} \sum_{p'} \tilde{W}(p, p') E^{V1}(p', t; \theta_i, v^c), \tag{7}$$

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

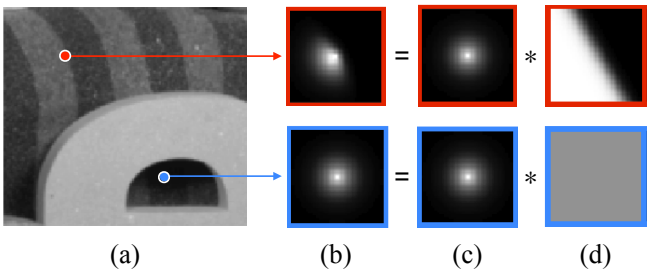

Fig. 1. Example of pooling weights at different positions: (a) Sample input indicating two different positions p (see red and blue dots) at which we show: (b) the final pooling weight $W(\cdot, p)$ which is obtained by multiplying (c) the isotropic term by the (d) anisotropic term (see text).

and where $\bar{N}(p, \theta_i) = \sum_{p'} \tilde{W}(p, p')$ 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}:

$$\alpha(\|R\|(p)) = \alpha_{max} e^{-\eta \frac{\|R\|^2(p)}{r_{max}}}, \tag{8}$$

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

$$g_i(p, p') = S_{\lambda,\nu}\left(-\frac{\nabla R_{\theta_i}(p)}{\|\nabla R_{\theta_i}\| + \varepsilon} \cdot (p' - p)\right), \tag{9}$$

where $S_{\lambda,\nu}(x) = 1/(1 + \exp(-\lambda(x - \nu)))$ 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. 1 gives two examples of the pooling coefficients at different positions.

3.3 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) whilst preserving discontinuities in motion and illumination. To do so, we propose an iterated trilateral filtering defined by:

$$u^{n+1}(p) = \frac{1}{\bar{N}(p)} \sum_{p'} W(p, p') u^n(p'), \tag{10}$$

$$c^{n+1}(p) = c^n(p) + \lambda(\max_{p' \in \mathcal{N}(p)} c^n(p') - c^n(p)) \tag{11}$$

$$u^0(p) = E^{MT}(p, t; \theta_i, v^c), \tag{12}$$

$$c^0(p) = C(p), \tag{13}$$

where

$$W(p,p') = c^n(p') f_\alpha(\|p-p'\|) f_\beta(c^n(p)(u^n(p') - u^n(p))) \\ f_\gamma(I(p') - I(p)) u^n(p'), \quad (14)$$

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.

4 Results

In order to test the proposed method, a coarse-to-fine 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 the Yosemite sequence (without clouds) as it is widely used in both computer vision and biological vision studies (see Fig. 2, first row). For FFV1MT we have AEE=3.55 ± 2.92, and for AMPD AAE=2.94 ± 2.00, where AAE is the average angular error (with associated standard deviations) [1]. This can be compared to what has been obtained with previous biologically-inspired models such as the original Heeger approach [2] (AAE=11.74°) and the neural model proposed in [3] (AAE=6.20°), 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°), showing a better performance.

The results on the Middlebury training set show improvements of the proposed method (see Table 1): in particular, AMPD improves the results of 18%, by considering the average AAE (aAAE) for all the sequences (aAAE=7.40°), with respect to FFV1MT (aAAE=9.05°). By considering state of the art computer vision approaches [16], our model (average EPE for all the sequences, aEPE=0.71 pixel) performs better than some algorithms, e.g. FlowNetC (aEPE=0.93 pixel), but other algorithms outperform it, e.g. SPyNet (aEPE=0.33 pixel).

For qualitative comparison, sample results are also presented in Fig.2. The relative performance can be understood by observing δAAE (last column of Fig. 2), difference between the FFV1MT AAE map and the AMPD AAE map: the improvements are prominent at the edges.

In order to assess the influence of the two cortical mechanisms (the V2-Modulated Pooling and the MT Lateral Interactions, see Section 3.2 and 3.3, respectively) on the optic flow computation, we have alternatively removed one of the two mechanisms from the AMPD model: the relative contribution of the V2-Modulated Pooling (aAAE=8.32°) and of the MT Lateral Interactions

	FFV1MT		AMPD	
Sequence	AAE ± STD	EPE ± STD	AAE ± STD	EPE ± STD
grove2	4.28 ± 10.25	0.29 ± 0.62	3.71 ± 8.95	0.25 ± 0.54
grove3	9.72 ± 19.34	1.13 ± 1.85	9.42 ± 18.41	1.00 ± 1.62
Hydrangea	5.96 ± 11.17	0.62 ± 0.96	5.83 ± 11.41	0.51 ± 0.71
RubberWhale	10.20 ± 17.67	0.34 ± 0.54	6.69 ± 10.92	0.24 ± 0.34
urban2	14.51 ± 21.02	1.46 ± 2.13	11.91 ± 18.98	1.01 ± 1.41
urban3	15.11 ± 35.28	1.88 ± 3.27	11.31 ± 29.73	1.24 ± 2.17

Table 1. Error measurements, AAE and EPE (endpoint error), on Middlebury training set

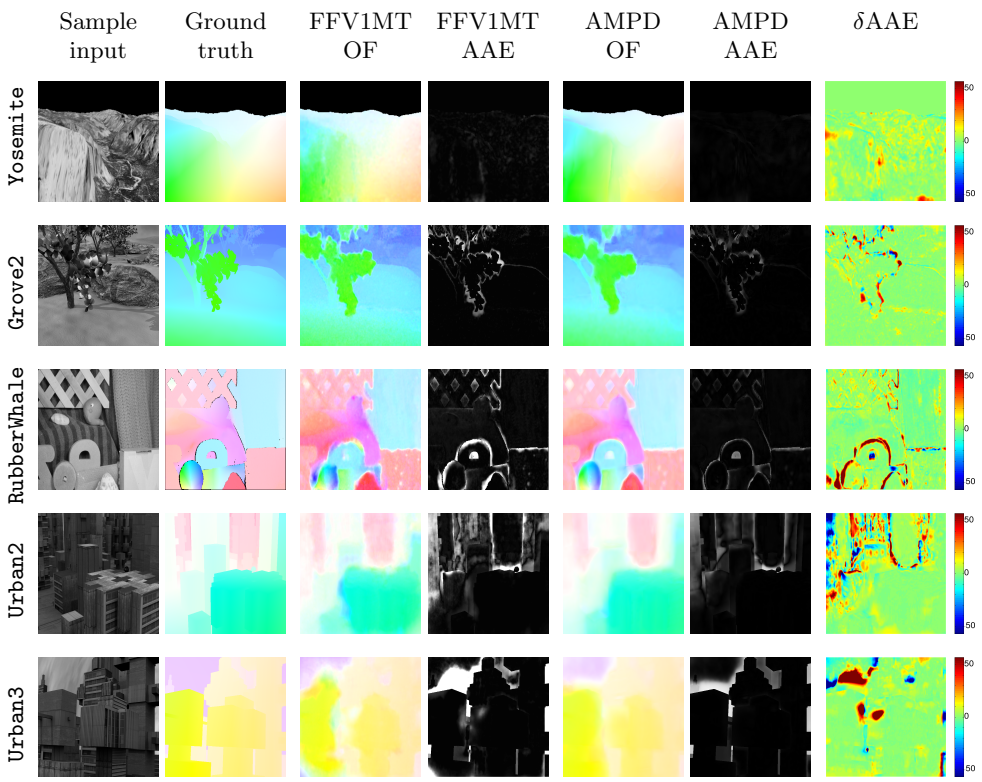

Fig. 2. Sample results on Yosemite sequence and a subset of Middlebury training set. δAAE = AAE_{FFV1MT} - AAE_{AMPD}

(aAAE=8.31°) is similar, which corresponds to an improvement of 8%. In order to qualitatively highlight the relative contribution of the different neural mechanisms on optic flow computation, Fig. 3 shows an enlarged region of the RubberWhale sequence. It is worth noting that the main effect of the two devised mechanisms is on borders and discontinuities.

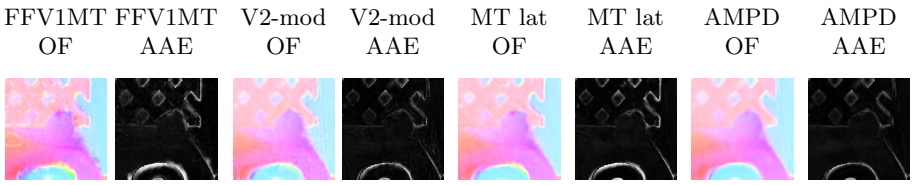

Fig. 3. Comparison of the effects on optic flow computation of the different neural mechanisms considered: in particular "V2-mod" refers to the V2-Modulated Pooling and "MT lat" to the MT Lateral Interactions.

5 Conclusion

In this paper, we have proposed the new brain-inspired algorithm AMPD that incorporates three functional principles observed in primate visual system, namely contrast adaptation, image structure based afferent pooling and ambiguity based lateral interaction. The AMPD is an extension of the state of the art algorithm FFV1MT [20], which is appreciated by both computer vision and biological vision communities. Contemporary computer vision methods to Heeger et al. [19], such as Lucas and Kanade and Horn and Schunck, which study local motion estimation and global constraints to solve aperture problem, have been revisited by the computer vision with great interest [6] 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. Very few attempts have been made to incorporate these ideas into spatio-temporal filter based models, and given the recent growth in neuroscience, it is very interesting to revisit this model incorporating the new findings and examining the efficacy. Differently from FFV1MT and Spynet [16], which only rely on scale space for diffusion of non-local cues, our AMPD model provides a clue on the potential role played by the recurrent interactions in solving the blank wall problem by non local cue propagation. It is also worth noting that bilateral filtering based techniques are gaining popularity in semantic segmentation using convolutional neural networks. Here, we show how neural modulation based on local context amounts to such bilateral filtering and a promising direction to explore even for dense optical flow.

The AMPD improves the flow estimation compared to FFV1MT and 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 RFs, 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.

References

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

2. Barron, J., Fleet, D., Beauchemin, S.: Performance of optical flow techniques. International Journal of Computer Vision 12(1), 43–77 (1994)
3. Bayerl, P., Neumann, H.: Disambiguating visual motion through contextual feedback modulation. Neural computation 16(10), 2041–2066 (2004)
4. Bouecke, J.D., Tlapale, E., Kornprobst, P., Neumann, H.: Neural mechanisms of motion detection, integration, and segregation: From biology to artificial image processing systems. EURASIP J. on Advances in Signal Processing 2011, 6 (2011)
5. Bradley, D.C., Goyal, M.S.: Velocity computation in the primate visual system. Nature Reviews Neuroscience 9(9), 686–695 (2008)
6. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. International Journal of Computer Vision 61(3), 211–231 (2005)
7. Butler, D.J., Wulff, J., Stanley, G.B., Black, M.J.: A naturalistic open source movie for optical flow evaluation. In: European Conference on Computer Vision. pp. 611–625. Springer (2012)
8. Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., Movshon, J.A.: A functional and perceptual signature of the second visual area in primates. Nature neuroscience 16(7), 974–981 (2013)
9. Ilg, U., Masson, G.: Dynamics of Visual Motion Processing: Neuronal, Behavioral, and Computational Approaches. Springer e-Books, Springer Verlag (2010)
10. Kovesi, P.: Image features from phase congruency. Videre: Journal of computer vision research 1(3), 1–26 (1999)
11. Maunsell, J.H., Van Essen, D.C.: 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)
12. Medathati, N.V.K., Neumann, H., Masson, G.S., Kornprobst, P.: Bio-inspired computer vision: Towards a synergistic approach of artificial and biological vision. Computer Vision and Image Understanding 150, 1 – 30 (2016)
13. Orban, G.A.: Higher order visual processing in macaque extrastriate cortex. Physiological reviews 88(1), 59–89 (2008)
14. Paninski, L.: Maximum likelihood estimation of cascade point-process neural encoding models. Network: Computation in Neural Systems 15(4), 243–262 (2004)
15. Rad, K.R., Paninski, L.: Information rates and optimal decoding in large neural populations. In: Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F.C.N., Weinberger, K.Q. (eds.) NIPS. pp. 846–854 (2011)
16. Ranjan, A., Black, M.: Optical flow estimation using a spatial pyramid network. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Jul 2017)
17. Rust, N.C., Mante, V., Simoncelli, E.P., Movshon, J.A.: How MT cells analyze the motion of visual patterns. Nature Neuroscience 9(11), 1421–1431 (2006)
18. Sharpee, T.O., Sugihara, H., Kurgansky, A.V., Rebrik, S.P., Stryker, M.P., Miller, K.D.: Adaptive filtering enhances information transmission in visual cortex. Nature 439(7079), 936–942 (2006)
19. Simoncelli, E.P., Heeger, D.J.: A model of neuronal responses in visual area MT. Vision research 38(5), 743–761 (1998)
20. Solari, F., Chessa, M., Medathati, N.K., Kornprobst, P.: What can we expect from a V1-MT feedforward architecture for optical flow estimation? Signal Processing: Image Communication 39, 342–354 (2015)