

What can we expect from a classical V1-MT feedforward architecture for optical flow estimation?

Fabio Solari, Manuela Chessa, Kartheek Medathati, Pierre Kornprobst

▶ **To cite this version:**

Fabio Solari, Manuela Chessa, Kartheek Medathati, Pierre Kornprobst. What can we expect from a classical V1-MT feedforward architecture for optical flow estimation?. [Research Report] RR-8618, INRIA Sophia Antipolis; University of Genoa - DIBRIS, Italy; INRIA. 2014, pp.22. hal-01078117

HAL Id: hal-01078117

https://inria.hal.science/hal-01078117

Submitted on 28 Oct 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

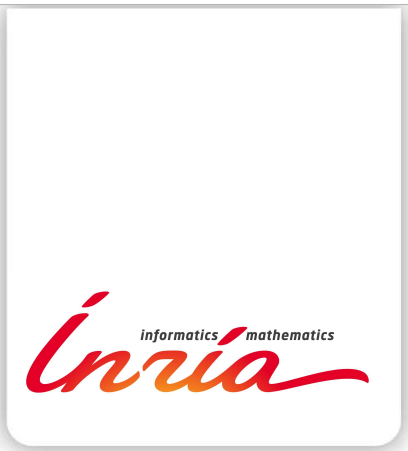


What can we expect from a classical V1-MT feedforward architecture for optical flow estimation?

Fabio Solari, Manuela Chessa, N V Kartheek Medathati, Pierre Kornprobst

RESEARCH REPORT
N° 8618
October 2014
Project-Team Neuromathcomp

ISSN 0249-6399 ISRN INRIA/RR--8618--FR+ENG



What can we expect from a classical V1-MT feedforward architecture for optical flow estimation?

Fabio Solari*, Manuela Chessa*, N V Kartheek Medathati†, Pierre Kornprobst†

Project-Team Neuromathcomp

Research Report n° 8618 — October 2014 — 23 pages

Abstract: Motion estimation has been studied extensively in neurosciences in the last two decades. The general consensus that has evolved from the studies in the primate vision is that it is done in a two stage process involving cortical areas V1 and MT in the brain. Spatio temporal filters are leading contenders in terms of models that capture the characteristics exhibited in these areas. Even though there are many models in the biological vision literature covering the optical flow estimation problem based on the spatio-temporal filters little is known in terms of their performance on the modern day computer vision datasets such as Middlebury. In this paper, we start from a mostly classical feedforward V1-MT model introducing a additional decoding step to obtain an optical flow estimation. Two extensions are also discussed using nonlinear filtering of the MT response for a better handling of discontinuities. One essential contribution of this paper is to show how a neural model can be adapted to deal with real sequences and it is here for the first time that such a neural model is benchmarked on the modern computer vision dataset Middlebury. Results are promising and suggest several possible improvements.

Key-words: Optical flow, spatio-temporal filters, motion energy, cortical areas V1 and MT, benchmarking, Middlebury dataset

* University of Genoa - DIBRIS, Italy
† Inria, Project-Team Neuromathcomp, France

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Que peut-on attendre d'une architecture feedforward classique de V1-MT pour estimer le flot optique?

Résumé : L'estimation de mouvement est un sujet largement étudié en neurosciences au cours des deux dernières décennies. Le consensus général qui a évolué à partir des études chez la vision des primates est que cela est fait dans un processus en deux étapes impliquant les aires corticales V1 et MT. Les filtres spatio-temporels sont en tête des prétendants en termes de modèles qui intègrent les caractéristiques trouvées dans ces domaines. Malgé les nombreux modèles dans la littérature de la vision biologique couvrant le problème de l'estimation du flot optique à partir de filtres spatio-temporels, nous en connaissant peu quant à leur performance sur des jeux de données modernes de vision par ordinateur comme celui de Middlebury. Dans ce papier, nous partons d'un modèle essentiellement classique des aires V1-MT en y introduisant une étape de décodage supplémentaire pour obtenir une estimation du flot optique. Deux extensions sont également examinées utilisant un filtrage non linéaire de la réponse MT pour un traitement des discontinuités. Une contribution essentielle de cet article est de montrer comment un modèle neuronal peut être adapté pour pouvoir être testé sur des séquences réelles et c'est la première fois qu'un tel modèle neuronal est évalué avec Middlebury. Les résultats sont prometteurs et suggèrent plusieurs améliorations possibles.

Mots-clés : Flot optique, filtres spatio-temporels, énergie de mouvement, aires corticales V1 et MT, évaluation, données de Middlebury

1 Introduction

Interpretation of visual motion information is a key competency for biological vision systems to survive in a dynamic world and also for computer vision as it could enable several applications. Owing to application potential, optical flow estimation which is a fundamental problem in visual motion analysis has been studied extensively by both computer vision and biological vision communities. Estimating optical flow refers to assignment of 2-D vectors at sample locations of the visual image that describe their displacements within the sensor's frame of reference using the change of structured light in the retinal or camera image. This displacement vector field constitutes the image flow representing apparent 2-D motions resultant of their 3-D velocities being projected onto the sensor. Such 2-D motions are observable only at intensity variations (and are thus contrast dependent) as a consequence of the relative change between an observer (eye or camera) and the surfaces or objects in a visual scene.

In the past two decades efforts by computer vision researchers have lead to development of a large number of algorithms for computation of optical flow with majority of them extracting flow field using two consecutive frames in the video sequence (see [1] for a review). Despite a large amount of work in this area, the problem remains still hard to solve as many of the algorithms either lack consistent accuracy across video sequences or suffer from computational cost. One of the prominent achievements of the research efforts that went in computer vision could be development of publicly available benchmarking datasets that provided an exciting opportunity to evaluate models and learn their short comings of the predictions in case of natural image scenarios.

On the other hand the neural mechanisms underlying motion analysis in the visual cortex have been extensively studied almost with little interaction with computer vision community [2, 3, 4, 5, 6, 7, 8, 9] resulting in few mathematical models. Even though there was some early interaction among the two communities [10, 11, 12] comparatively little work has been done in terms of examining or extending the mathematical models proposed in biology in terms of their engineering efficacy on modern optical flow estimation datasets.

In this work we take a step towards filling the critical gap between biological and computer vision communities working on visual motion estimation leveraging and testing ideas proposed in biology in terms of building scalable algorithms. This is a challenging task as many of the mathematical models proposed in biology are confined to highly primed stimuli or often only examine a local decision making process such as a receptive field (RF) property which demands non-trivial extensions to be made before the ideas could be tested on the complex real world datasets.

The paper is organized as follows: In Sec 2 we begin by presenting a brief summary of the models describing biological visual motion estimation with emphasis on spatio-temporal filter based ones, and present our **baseline model** which is a three-step feedforward model. The two first steps correspond to V1 and MT pattern cells responses and follow classical ideas from the literature. The third step is a decoding stage to obtain the optical flow from MT population response. Then we propose **extensions** of the baseline model by introducing a filtering stage at the level of the MT pattern cell layer in order to better preserve motion discontinuities. In Sec. 3 we present the algorithmic details which are an essential contribution here since it is the first time that this kind of V1-MT architecture is evaluated on modern computer vision datasets. We present a multi scale processing strategy to deal with a large range of displacements and the solutions to deal with boundary conditions and unreliable regions. In Sec. 4, we evaluate our approach, and parameters setting is described. First we use some test sequences to show the intrinsic properties of our approach and the influence of some parameters. Then we benchmark our approach using the modern computer vision datasets Middlebury [1]. Finally, we present the

conclusion of our work in Sec. 5.

2 Model description

2.1 Context: Biological visual motion estimation

Two primary cortical areas involved in motion estimation are V1 [13, 14] and MT [15] (for reviews see, e.g., [16, 17, 18]).

Several models have been devised in order to account for the neural mechanisms of motion analysis. A model for the extraction of the image flow, inspired by the stages of the visual motion pathway of the primate brain, is proposed in [19]. Such a model is based on spatio-temporal RFs of simple cells, modeled by Gabor functions [20], and on the computation of the motion energy [21] by a layer of complex cells. The problem of detecting local image velocity is also addressed in [22], by a two-layer model that first computes motion energies and eventually estimates velocity by combining these energies. The combination of the responses of the different cells, necessary to solve the aperture problem and to compute velocity, is the main difference between [19] and [22]. More recently, computational models to compute local image velocity through a two-layer architecture corresponding to V1 and MT cortical areas have been proposed [12, 5].

In general, the pattern selectivity of MT cells can be explained by following two different approaches [17, 18]. In [17], the authors consider the motion computation related to some kind of 2-D feature extraction mechanism. The consequence is that the aperture problem does not affect the motion processing, though few evidence for a feature-tracking mechanism are reported [23, 24, 25]. In [18] which our neural model is based on, the authors use a non-linear integration of the V1 afferents to obtain the MT pattern cells: in particular, the intersection of constraints (IOC) mechanism is indirectly considered through localized activations of V1 cells [26, 12, 5].

However, this class of models, although conceptually valuable, has always been characterized by a high computational cost, as the price to pay for mimicking the hierarchical processing of the visual signal. Indeed, these kinds of neural solutions were developed as models of how the visual cortex works, and their functionality were demonstrated on simple synthetic sequences, but were never designed to form a systematic alternative to computer vision algorithms, working on real video sequences. Here, it is the first time that a neural model is chosen as a starting point and adapted so that it could be tested against modern computer vision benchmarks.

2.2 Model overview

A global overview of the model studied in this paper is given in Fig. 1. It is a three-step feedforward model: Step 1 corresponds to the V1 simple and complex cells, Step 2 corresponds to the MT pattern cells and Step 3 corresponds to a decoding stage to obtain the optical flow from the MT population response. In term of modeling, Steps 1 and 2 follow a classical view, while Step 3 has been introduced to solve the task of optical flow.

Let us 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. Details for each steps are given in the following sections.

2.2.1 Step 1 : V1 (*Motion energy estimation and normalization*)

In the V1-layer two sub-populations of neurons are involved in the information processing, namely V1-direction selective simple cells and complex cells.

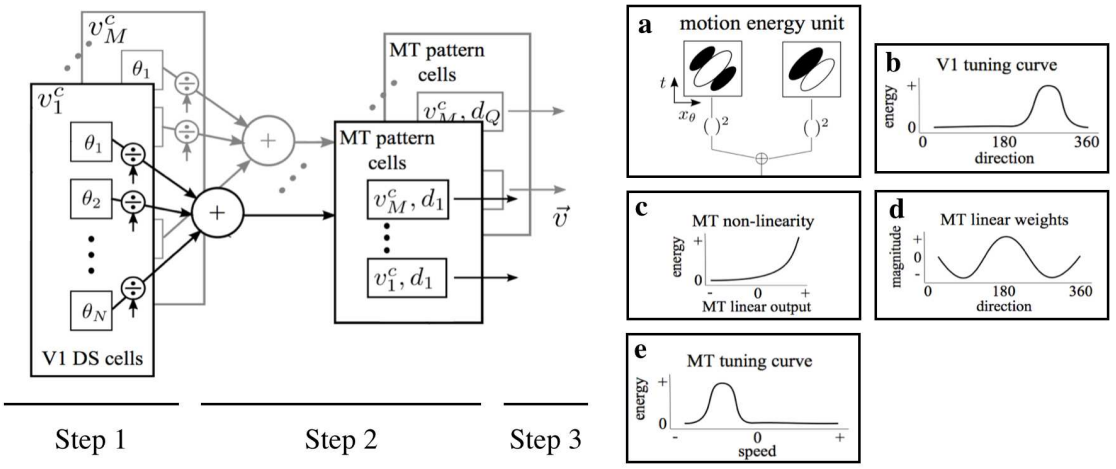


Figure 1:
Baseline model overview: It is a three-step feedforward model where the first step corresponds to the V1 layer (motion energy estimation and normalization), the second step corresponds to an MT layer (motion energy pooling and non-linearity) and the third step is velocity estimation (decoding). Insets show different properties of each step.

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 p. 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:

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

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

where σ and τ are define the spatial and temporal scales respectively. Denoting the real and imaginary components of the complex filters h and p as h_e, p_e and h_o, p_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,

$$g_o(x, y, t; \theta, v^c) = h_o(x, y; \theta, f_s)p_e(t; f_t) + h_e(x, y; \theta, f_s)p_o(t; f_t),$$
$$g_e(x, y, t; \theta, v^c) = h_e(x, y; \theta, f_s)p_e(t; f_t) - h_o(x, y; \theta, f_s)p_o(t; f_t).$$

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

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

The complex cells are described as a combination of the quadrature pair of simple cells (4) by using the motion energy formulation (Fig. 1, inset a),

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

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

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

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 (Fig. 1, inset b), since the preferred velocity of each cell is related to the direction orthogonal to its spatial orientation.

2.2.2 Step 2: 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 [12]. The responses of an MT pattern cell tuned to the speed v^c and to direction of speed d can be expressed as follows:

$$E^{MT}(x,y,t;d,v^c) = F\left(\sum_{i=1}^{N} w_d(\theta_i) G_{\sigma_{pool}} \overset{x,y}{*} E^{V1}(x,y,t;\theta_i,v^c)\right), \tag{6}$$

where $G_{\sigma_{pool}}$ denotes a Gaussian kernel of standart deviation σ_{pool} for the spatial pooling on the neighboring locations, $F(s) = exp(s)$ is a static nonlinearity chosen as an exponential function [27, 5] (Fig. 1, inset c), and w_d represents the MT linear weights that give origin to the MT tuning (Fig. 1, inset d). 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 [28]. Considering the MT linear weights shown in [5], $w_d(\theta)$ is defined by θ:

$$w_d(\theta) = \cos(d-\theta) \quad d \in [0, 2\pi[. \tag{7}$$

This choice allows us to obtain direction tuning curves of pattern cells that behave as in [5] (see illustration in Fig. 1, inset e).

2.2.3 Step 3: Velocity estimation

Given the target of this approach to be an estimation of dense optical flow, the velocity vector has to be obtained by decoding the population responses of the MT neurons. Indeed, a unique velocity vector cannot be recovered by activity of a single velocity tuned MT neuron as multiple scenarios could evoke the same activity, but unique vector can be recovered based on the activity of a population. In this section, we present a decoding step which was not present in [12, 5] 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 [29].

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

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

2.3 Extended approaches

As it will be shown in Sec. 4, the results obtained with the baseline model suffer from two main problems which are a direct consequence of the model choice. Resulting optical flow can be noisy in areas with smooth speeds, and discontinuities are lost since there is no mechanism in the baseline model to tackle these issues.

To reach a better performance, we propose to introduce a nonlinear filtering stage at the level of MT responses. Two kinds of informations can be taken into account to smooth the resulting optical flow while preserving its discontinuities. The first is directly based on the MT response and the idea is to prevent diffusion where MT response appears to be discontinuous in space. The second takes into account luminance similarity to gate diffusion, since in most real world sequences, motions discontinuities are observed at objects boundaries (and thus luminance discontinuities). Here, we propose solutions based on classical non-linear filters used in computer vision: bilateral or trilateral filtering. Note that these solutions do not have any direct biological interpretation. Although some analogies could be thought in term of functional principles with motion contrast cells in MT and some V2/V3 neurons, it is beyond the scope of the paper to give a truly and rigorous bio-inspired modeling of these mechanisms.

2.3.1 Bilateral filter solution (BF)

The first solution is based on the bilateral filter (BF) which can be interpreted as an extension of the linear Gaussian filtering. This method has been extensively used in the context of image smoothing leading to many applications (see [32] for a review). Given an image to be smoothed, the idea of BF is to consider that two pixels are close to each other not only if they occupy nearby spatial locations but also if they have some similarity in the values range. As a consequence, the interest of BF is that it allows to smooth an image while keeping its discontinuities.

In our context, we propose to apply BF on $E^{MT}(x,y,t;d,v^c)$ responses (at each time t and for each d, v^c independently). Denoting $E^{MT}(x,y,t;d,v^c)$ by $E^{MT}(p)$ where $p = (x,y)$ for sake of simplicity, the bilateral filter is defined by

$$BF_{\alpha,\beta}E^{MT}(p) = \frac{1}{N(p)} \int_{p' \in \Omega} f_\alpha(\|p - p'\|) f_\beta(E^{MT}(p') - E^{MT}(p)) E^{MT}(p') dp', \tag{9}$$

where α and β are two parameters determining the smoothing property and

$$f_\mu(s) = \exp(s^2/\mu^2) \quad s \in R, \tag{10}$$

and $N(p)$ is the normalizing term

$$N(p) = \int_{p' \in \Omega} f_\alpha(\|p - p'\|) f_\beta(E^{MT}(p') - E^{MT}(p)) dp',$$

Equation (9) can be interpreted as a nonlinear weighted sum of the values of E^{MT} taken at positions p' close to p and such that values $E^{MT}(p')$ are also close to $E^{MT}(p)$. By doing so, the resulting filtered energy $BF_{\alpha,\beta}E^{MT}$ is smoothed while main discontinuities are preserved and enhanced. Several iterations of this filter can be made depending on the degree of smoothing desired.

2.3.2 Trilateral filter solution (TF)

A direct extension of (9) is to add an extra weighting term which depends on luminance similarity $f_\gamma(I(p') - I(p))$. The proposed trilateral filter is then defined by

$$TF_{\alpha,\beta,\gamma}E^{MT}(p) = \frac{1}{N(p)} \int_{p' \in \Omega} f_\alpha(\|p - p'\|) f_\beta(E^{MT}(p') - E^{MT}(p)) f_\gamma(I(p') - I(p)) E^{MT}(p') dp', \tag{11}$$

where α, β, γ are parameters and $N(p)$ is the normalizing term

$$N(p) = \int_{p' \in \Omega} f_\alpha(\|p - p'\|) f_\beta(E^{MT}(p') - E^{MT}(p)) f_\gamma(I(p') - I(p)) dp'.$$

The result of Eq. (11) is that the difference of luminance between $I(p)$ and $I(p')$ will be also taken into account when considering $E^{MT}(p')$. This luminance gated mechanism have some similarities with the former bio-inspired models [7, 33] although modeled differently here.

3 Algorithmic details: How to make the approach applicable to real sequences

The classic V1-MT model presented in Sect. 2 is primarily aimed at explaining the recorded neural and perceptual data on largely homogeneous synthetic images such as moving gratings and plaids. As our objective is to benchmark it on real sequences, we need to introduce algorithmic solutions to address several problems coming from the nature of real sequences but also from our model.

3.1 Multiscale approach

As detailed in the Section 2, the V1-like RFs are modeled through spatio-temporal filters. In order to keep as low as possible the computational load of the model, only one spatial radial peak frequency f_s has been considered. This is in contrast with the physiological findings, and since information in natural images is spread over a wide range of frequencies, it is necessary to use a mechanism that allows us to get information from the whole range.

In this paper, a multi-scale processing (i.e., considering units tuned to different spatial peak frequencies) is implemented (Fig. 2). This is a classical approach used in computer vision. It consists in (i) a pyramidal decomposition [34] with L levels and (ii) a coarse-to-fine refinement [35], which is a computationally efficient way to take into account the presence of different spatial frequency channels in the visual cortex and their interactions. The optical flow at a finer scale

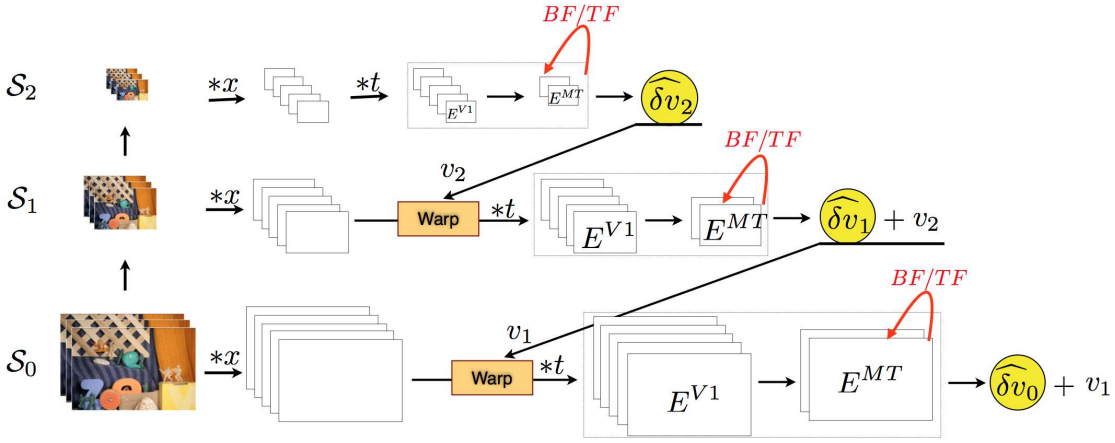


Figure 2: How the multi-scale processing works: Starting from the pyramidal decomposition $\mathcal{S}_l (l = 0 \dots L-1)$ ($l = 0$ is the finer scale), the velocities obtained at a coarser level of the pyramid ($\widehat{\delta v_l}$), are expanded (v_l) and used to warp the sequence of the spatially filtered images, then the residual optic flow is computed ($\widehat{\delta v_{l-1}}$).

($\mathcal{S}_l$) is given by $v_l = \widehat{\delta v_l} + v_{l+1}$. Using this approach, the spatial distance between corresponding points is reduced, thus yielding to a more precise estimate, since the residual values of the velocities lie in the filters' range. This also allows large displacements to be estimated which is a crucial aspect when dealing with real sequences. Interestingly, at a functional level, there is an experimental evidence that MT neurons seems to follow a coarse-to-fine strategy [3] suggesting that motion signals become more refined over time.

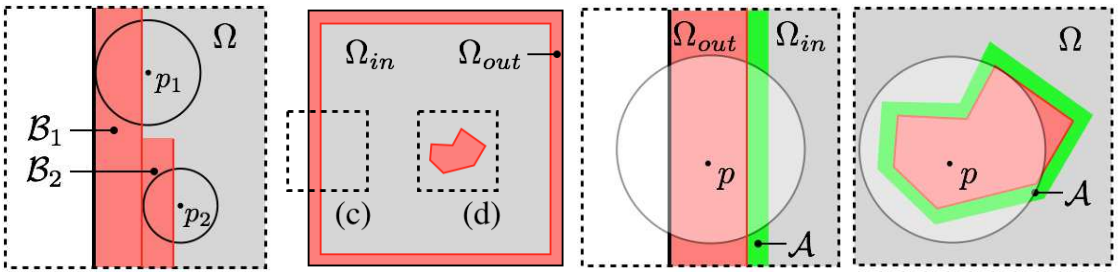
3.2 Boundary conditions

The problem of boundary conditions arises as soon as we need to consider values outside the domain of definition Ω. Even with a simple Gaussian smoothing, when estimating results close to the boundaries, one needs to access values outside Ω. This is solved generally by choosing some boundary conditions like Neumann or Dirichlet. However, in our case, using such assumptions might introduce some strong errors at the boundaries. For this reason, we proposed instead to work on an inner region in which only available values are taken into account so that no approximation or assumption has to be made, and then to interpolate values in the remaining outer region. Note that this is an important issue to consider, especially because we use a multi-scale approach since errors done at the boundaries at low scales can spread a lot as scales is getting finer.

The elements allowing to define the inner region Ω_{in} are illustrated in Fig. 3(a). Region $\mathcal{B}_1$ is first excluded so that V1 cell responses use only given values. Region $\mathcal{B}_2$ is then excluded because of the pooling operation from V1 to MT. Given this definition of inner and outer regions (Fig. 3(b)) the idea is to make all the estimations in Ω_{in} and to interpolate values in the outer region Ω_{out} (Fig. 3(c)). So, given E^{MT} estimated in Ω_{in}, we propose that

$$E^{MT}(p) = \frac{1}{N(p)} \int_{p' \in \mathcal{A}} f_\alpha(\|p - p'\|) f_\gamma(I(p) - I(p')) E^{MT}(p') dp' \quad \forall p \in \Omega_{out}, \tag{12}$$

where $\mathcal{A}$ contains pixels at the boundary and inside Ω_{in} (green region), function f_μ is defined as



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

Figure 3: Illustration of the filling-in approach used to deal with boundary conditions and the unreliable regions. (a) How inner domain is defined taking into account V1 filter spatial size and V1 to MT pooling; (b) Image domain showing the inner region Ω_{in} in grey where exact calculations (i.e., without any approximation) can be done, the outer region Ω_{out} where an interpolation scheme is chosen, and an example of unreliable region explained in (d); (c) Illustration of the interpolation scheme for a pixel $p \in \Omega_{out}$, showing the spatial neighbourhood associated with the spatial support of the integration and in green the region $\mathcal{A}$ which is used to estimate the interpolated values; (d) Same as (c) but in the case of an unreliable region.

in (10), α and γ are parameters and $N(p)$ is a normalising term

$$N(p) = \int_{p' \in \mathcal{A}} f_\alpha(\|p - p'\|) f_\gamma(I(p) - I(p'))dp'.$$

This method is based on luminance similarities, using the same idea as developed in Sec. 2.3. Note that other methods could be proposed such as more elaborated interpolation methods (like PDE-based inpainting methods) to further improve the quality of the results.

3.3 Unreliable regions

A problem is found with regions having a null spatio-temporal content, which is happening in the blank wall problem for example. In that case, locally, it is then not possible to find a velocity. Given a threshold T, a pixel (x, y) will be categorised as unreliable if and only if $E^{MT}(x, y, t; d, v^c) < T$ for all d and v^c. For these pixels, the same interpolation as (12) is proposed (Fig. 3(d)).

4 Results

4.1 Parameters settings

Table 1 provides parameters used in our simulations.

Let us discuss some of the main parameters. The choice of the size of the spatial support of the RF is related to the necessity of processing fine details, i.e. at high image resolution, in real-world sequences. The temporal support of the filter is due to two different needs: to obtain good results on the benchmarking sequences, and to meet the experimental evidence. In particular, the V1 and MT RFs process the visual signal within a restricted epoch of time, we can consider an average period of 200 ms [36, 3]. Such a time period corresponds to about 5 frames for a standard video acquisition device. Such choices allows us to obtain good performances in terms of the execution time, though they do not allow us to have tuning to higher velocities than 1

pixel/frame (i.e. the sampling constraints). The issues is addressed by considering a multi-scale approach, as explained in Sec. 3.1, the number of the considered scales depends on the size of the input images (and also on the speed range, though it is not a priori known), for the Middlebury sequences it has been fixed to 6 spatial scales. It is worth noting that to avoid the introduction of a loss of balance between the convolutions with the even and odd Gabor filters, the contribution of the DC component is removed [37]. Finally, the support of the spatial pooling $G_{\sigma_{pool}}$ is chosen in order to take into account the larger RFs of MT layer with respect to V1 layer, in particular the factor is 5, which is in accordance with findings reported in literature [38, 39].

All results presented here could be reproduced using our code available publicly.[1]

Description	Parameter	Value	Equation
V1			
RF spatial scale	σ	2.27	(1)
... and spatial support	SS	11×11 pixels,	(1)
Time constant of the exp. decay	τ	2.5	(2)
... and temporal support	TS	5 frames	(2)
Spatial radial peak frequency	f_s	0.25 cycles/pixel	(1)
Temporal radial peak frequencies	f_t	$\{0, 0.10, 0.15, 0.23\}$ cycles/frame	(3)
Number of spatial contrast orientations	N	8 (from 0 to π)	(5)
... and sampling	θ_i	$\theta = k\pi/N, k = 0..N-1$	(5)
Number of component speeds	M	7	(3)
... and sampling	v^c	$\{-0.9, -0.6, -0.4, 0, 0.4, 0.6, 0.9\}$	(3)
Semi-saturation constant	ε	10^{-9}	(5)
MT			
Std dev of the Gaussian spatial pooling	σ_{pool}	0.9	(6)
... and spatial support		5×5 pixels	(6)
Decoding step			
Number of MT direction tuning directions		2	(8)
... and sampling	d	$\{0, \pi/2\}$	(8)
Extended approaches			
Spatial parameter of BF	α	$\{0.50, 0.83, 1.16, 1.50, 1.83\}$ as a function of spatial scale	(9)
Range parameter of BF	β	1/6 of energy range	(9)
Luminance parameter of TF	γ	1/6 of luminance range	(11)
Algorithm			
Number of scales	L	6	
Spatial parameter of BF (interpolation)	α	2.5 pixels	(9)
Luminance parameter of BF (interpolation)	γ	1/6 of luminance range	(11)

Table 1: Parameter values used in our simulations. Equation number refers to the equation where it has been first introduced.

4.2 Analysis of proposed approaches

In this section, we evaluate the baseline model and it extensions referred by BF (Sec. 2.3.1) and TF (Sec. 2.3.1), by using some test sequences to show the intrinsic properties of our approach and to show the influence of some parameters. When ground truth optic flow is available, average angular error (AAE) and endpoint error (EPE) will be estimated (with associated standard deviations) [1].

[1] The link will be given, upon acceptance of the paper.

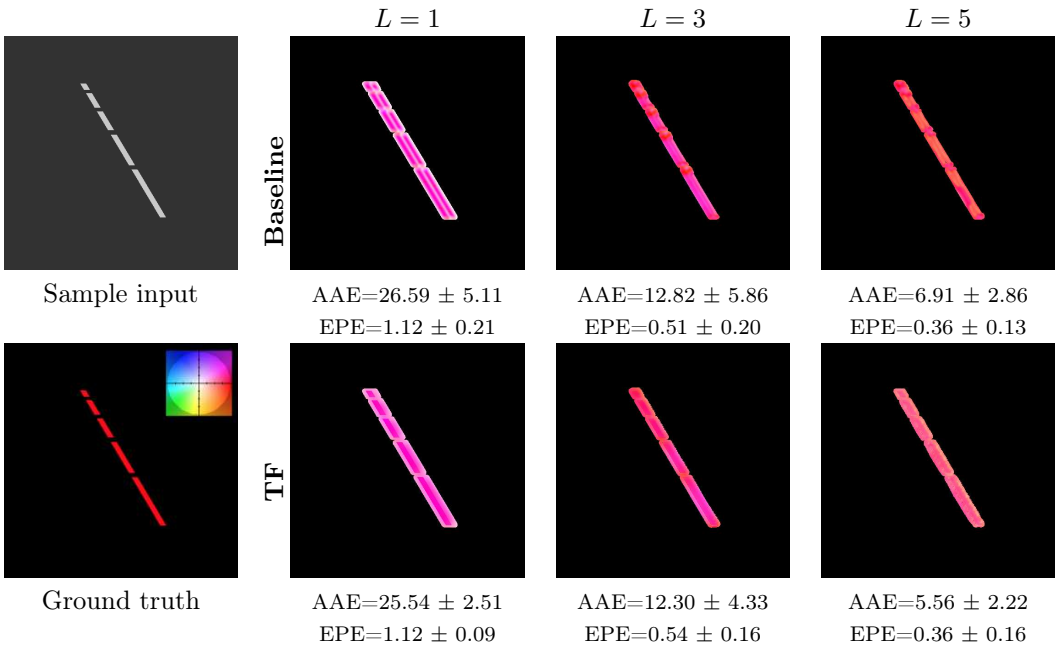


Figure 4: Influence of the number of scales. Baseline and TF approaches are tested with L=1, 3 and 5 scales.

The influence of the number of spatial scales is shown in Fig. 4. In this sequence a dashed bar moves rightward with velocity (2,0) pixels. Focusing first on the baseline method, results show that increasing the number of scales improves the results. It is worth noting that the aperture problem is correctly solved by considering 3 spatial scales in the small segments, whereas 5 spatial scales are needed to handle longer segments, though a residual optic flow at the finest scale is not correctly recovered in the middle of the longest segment, since the spatial support of the RFs is too small with respect to the visual feature. This issue is partially solved by the trilateral filter, which has the effect of diffusing further the estimates along the bar, which slightly improves the results.

The effects of the bilateral and of the trilateral filter applied at the MT level, with respect to the baseline approach, are further shown in Fig. 5. To do so, we considered a synthetic sequence that represents a textured translating shape moving with velocity (-3,-3) pixels on top of a translating background with velocity (4,0) pixels. Two situations are tested: In input 1, there is no brightness difference between background and foreground (only motion can reveal the object); In input 2 there a brightness difference. Ground truth for both are the same. Optical flows and error measurements confirm the improvement in optic flow estimation by considering the extended approach (in particular TF). The improvement is higher in the case of input 2 where the boundaries of the foreground shape are better preserved. Nevertheless, it is worth noting that more reliable optic flow estimates in the inner part of the background are obtained when no brightness differences are present. This is due to the fact that the multi-scale approach might create artefacts at low-resolution scales, in particular on sharp boundaries. Such artefacts produce wrong optic flow estimates at coarser spatial scales, which propagate at finer spatial scales through the warping stage.

In order to better analyze the roles of the different stages of the model, Fig. 6 shows the V1 and MT activities, by considering the synthetic sequence shown in Fig. 5. For V1, two maps are

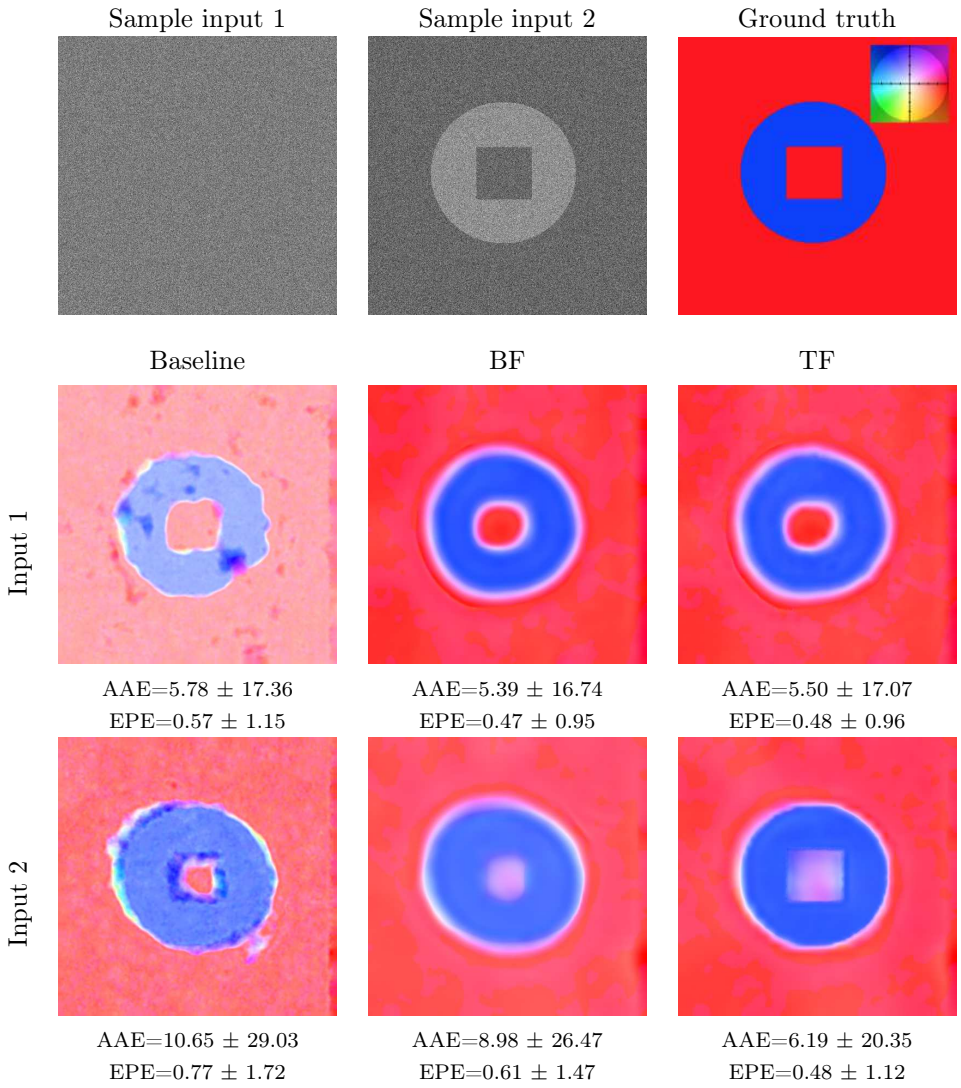


Figure 5: Comparison between the baseline, BF and TF approaches on a synthetic scene (a translating shape (-3,-3) pixels moving on top of a translating background (4,0) pixels).

shown:

$$\|E^{V1}\|_{\theta}(x,y,v^c) = \left(\sum_{i=1}^{N} E^{V1}(x,y;\theta_i,v^c)^2\right)^{1/2},$$

$$\|E^{V1}\|_{v^c}(x,y,\theta) = \left(\sum_{i=1}^{M} E^{V1}(x,y;\theta,v_i^c)^2\right)^{1/2},$$

and for MT, the two populations $E^{MT}(x,y,v^c,0)$ and $E^{MT}(x,y,v^c,\pi/2)$ are shown for different v^c. These results confirm show that the V1 layer has a tuning on the spatial orientation (thus the cells are elicited by the spatial orientation of the shape), whereas at MT layer, a speed tuning

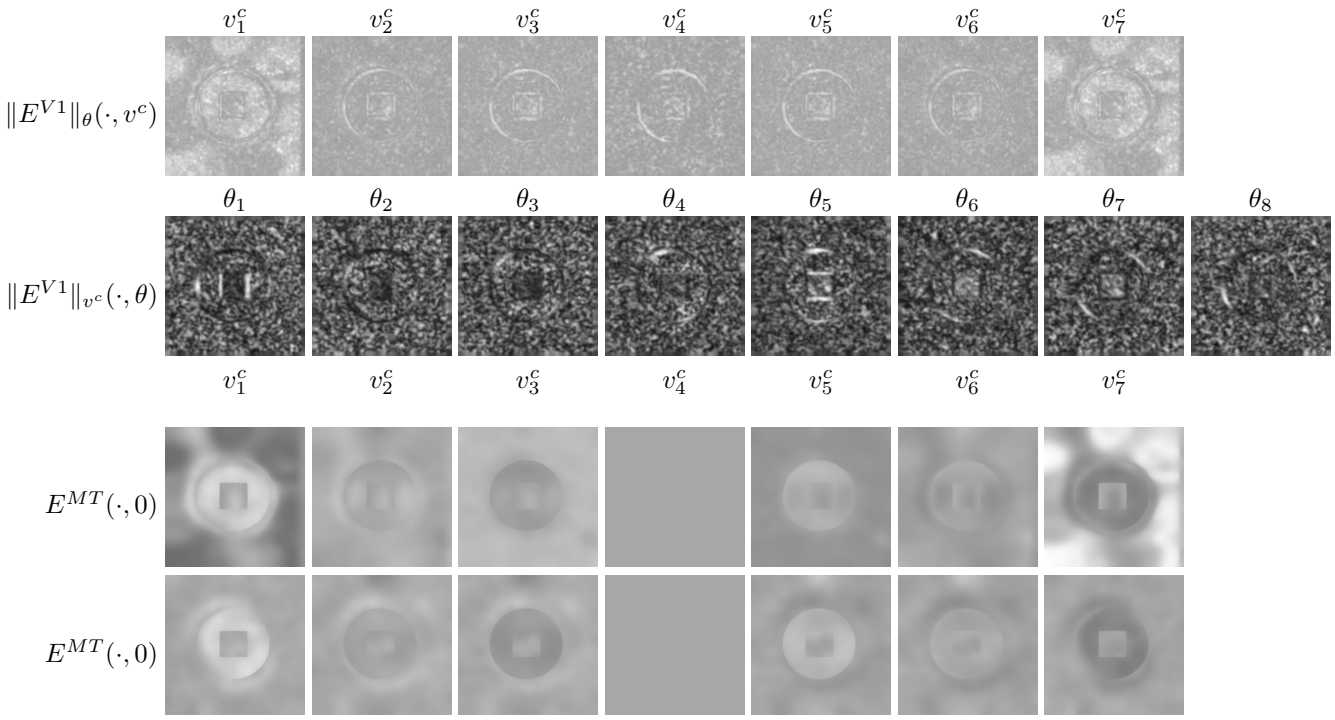


Figure 6: Illustrations of V1 and MT activities on the synthetic case shown in Fig. 5. The first row shows $\|E^{V1}\|_\theta(x, y, v^c)$: the activities do not identify specific tuning speeds, since all the spatial orientations are pooled in the norm and the tuning speeds are component speeds, i.e. they are orthogonal to the spatial orientation of the cell. The second row shows $\|E^{V1}\|_{v^c}(x, y, \theta)$: the cells are elicited by the spatial orientation of the shape, the V1 layer shows a tuning on the spatial orientation. The third and fourth rows show $E^{MT}(x, y, v^c, 0)$ and $E^{MT}(x, y, v^c, \pi/2)$ maps, respectively. At MT layer, a speed tuning emerges: on the left, the energies are higher for the region related to the shape, this means that there is a negative speed for the horizontal and vertical velocities related to the shape. On the right, the energies are higher for the background (for the third row, only)), since the background moves rightwards.

no more related to spatial orientation emerges (i.e., the aperture problem is solved).

Another representation consists in showing the distribution of E^{MT} at different positions to understand its relation to velocities (Fig. 7). By observing the behaviour of MT energies in four different positions on the frame (highlighted by (a), (b), (c) and (d) in the figure), it is possible to note how the MT layer encodes the velocities. In particular: the behaviours in (a) and (c) are affected by the values of the neighboring borders, thus there are no prominent activities; in (b), which corresponds to a point on the foreground shape, cells tuned to negative speeds (v_1^c) on both horizontal and vertical direction (E^{MT} with $d = 0$ and $d = \pi/2$, respectively) have the maximum response; in (d), which corresponds to a point on the background, only the response of the horizontal direction has a maximum for positive horizontal speed (v_7^c).

Finally we give a comparison between the baseline, BF and TF approaches on the classical realistic Yosemite sequence, without and with clouds (see Fig. 8). Both the maps representing computed optic flows and error measurements show a consistent improvement using TF: Optical flow is smoother with less artefacts. The average angular error on the sequence without clouds,

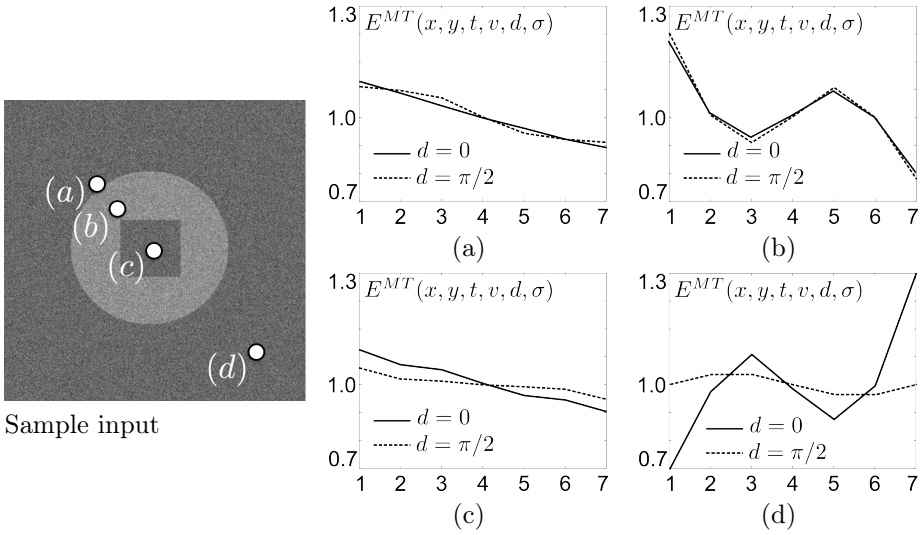


Figure 7: Distribution of MT energy at positions indicated in the sample input image

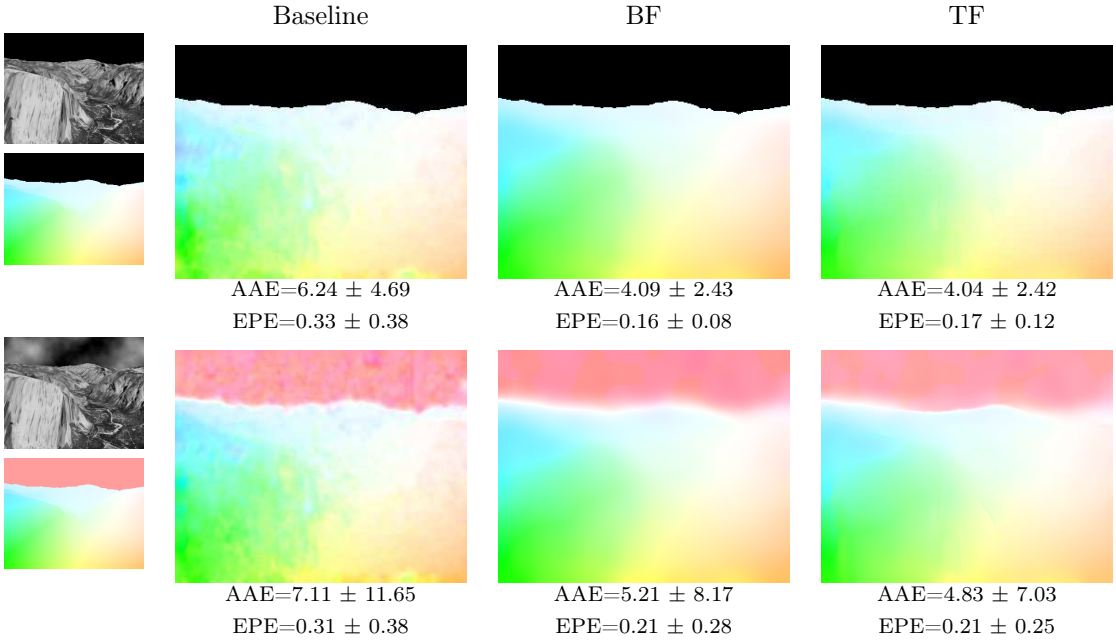


Figure 8: Comparison between the baseline, BF and TF approaches on the classical realistic `Yosemite` sequence, without and with clouds.

by considering the extended approach, is comparable with many of the approaches present in Middlebury evaluation table[2]. In the sequence with clouds, one can also notice the better preservation of discontinuties at some locations between the mountain and the sky.

[2]http://vision.middlebury.edu/flow/eval/results/results-a1.php

Sequence	AAE ± STD	EPE ± STD
grove2	4.72 ± 10.02	0.33 ± 0.62
grove3	10.65 ± 19.09	1.40 ± 1.92
Hydrangea	6.23 ± 11.25	0.80 ± 0.99
RubberWhale	8.18 ± 14.02	0.28 ± 0.44
urban2	13.37 ± 19.48	1.42 ± 1.45
urban3	18.82 ± 40.21	2.01 ± 3.10

Table 2: Error measurements on Middlebury training set

4.3 Performance evaluation on Middlebury dataset

In this Section, we benchmark our approach by using the modern computer vision datasets Middlebury [1][3]. In Fig. 9 and Table 2, we show qualitative and quantitative results obtained on the training dataset, which has public available ground truth. The sequences of this dataset have several challenges, such as sharp edges, high velocities and occlusions. Our model could fail in these situations essentially due to the support of the temporal filter: A small support of the temporal filters constrains the maximum detectable component velocity at each scale, thus the multi-scale approach is necessary in order to estimate higher velocities. On the other hand, the multi-scale approach creates artefacts at sharp edges, though the trilateral filter mitigates this issue. A large support of the temporal filters creates problems in occlusion regions.

Since `Urban2` sequence is quite challenging, due to the presence of both occlusions and high velocities, we decided to consider it, in order further analyze the proposed approach by varying some parameters of the model. In particular, we tested the effects of the size of the temporal support TS, the temporal decay τ and the number of tuning speeds. Figure 10 shows how the choice of these parameters affects the optical flow computation. The last column corresponds to the parameters chosen for the final evaluation on the Middlebury test set.

The optic flows computed on the evaluation Datasets of Middlebury have been submitted for evaluation [4]. Figure 11 shows the computed optic flow and the error maps, from which it is possible to note that higher errors are in correspondence of occlusions (wee, e.g., `Urban` sequence) and sharp edges (see, e.g., `Urban` and `Wooden` sequences). A screenshot of the Middlebury table is presented in Fig 12, from which we can see how our approach performs compared to other state-of-the-art algorithms. It is worth noting that our V1-MT-FF model is the only neural model for motion estimation present in the table.

5 Conclusion

In this paper, we have presented a method that is based on mathematical model primarily developed to account for various physiological findings related to motion processing in primates. Starting from the classical hierarchical feedforward processing model involving V1 and MT cortical areas which is usually limited to a single spatial scale, we have extended it to consider the whole visual field by adapting a multi scale approach and analyzed the efficacy of the approach in estimating the dense optical flow in real world scenarios by considering an efficient velocity decoding step.

We have tested the performance of our model using various synthetic stimuli as well as the standard Middlebury database. A qualitative evaluation indicates that model could recover

[3] http://vision.middlebury.edu/flow/data/

[4] http://vision.middlebury.edu/flow/eval/results-manuela-chessa/results-a1.php

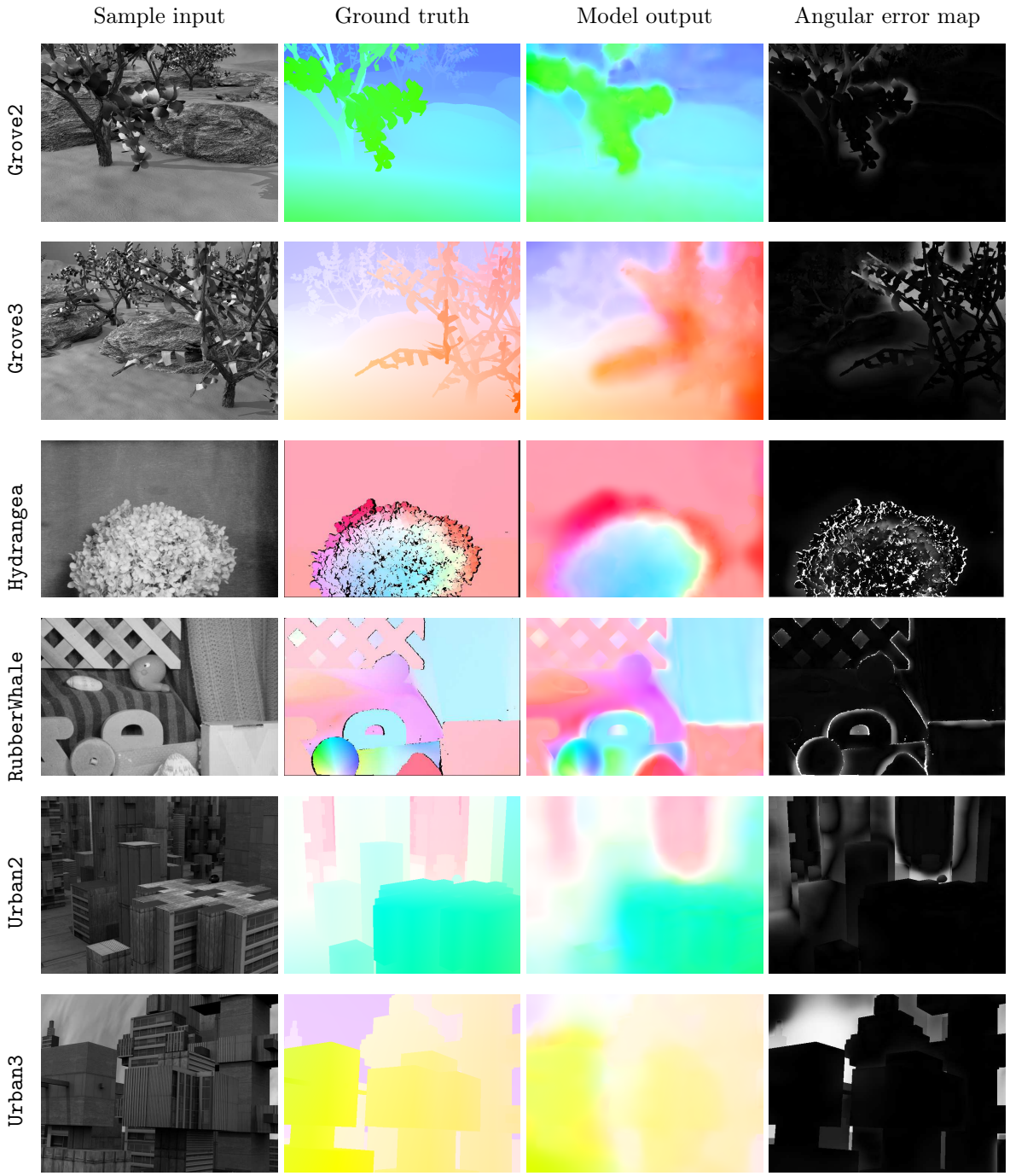


Figure 9: Sample results on Middlebury training set

velocity vectors in regions with coarse textures quite well, but typically fails to achieve robust estimates in regions with very fine texture or regions with sharp edges. This is to be expected,

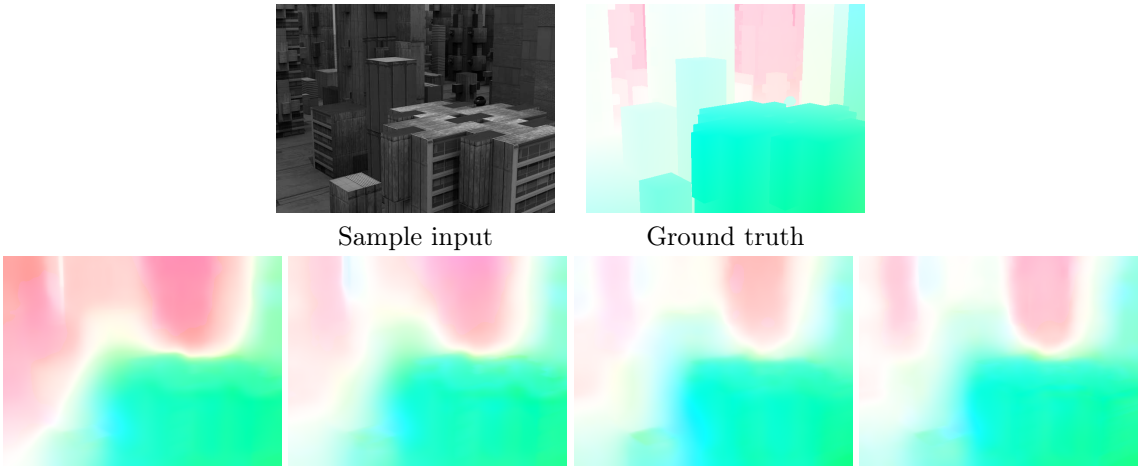


Figure 10: Effects of the size of the temporal support TS, the temporal decay τ and the number of tuning speeds on optical flow accuracy using `Urban2` sequence

since the V1-MT feedforward model does not take into account the details of lateral interactions and scale space issues that need to be tackled in order to solve the blank wall problem. In order to address this issue, we have incorporated as an extension to the baseline approach bilateral and trilateral filtering, which could potentially simulate lateral interactions at the MT level in order to obtain better dense estimations. Incorporation of trilateral filtering has significantly improved the estimation of dense optic flow maps, which reinforces the need to consider lateral interactions and possibly feedback into the models to better account for natural image processing scenarios, going beyond the homogeneous synthetic stimuli such as grating and plaids that are traditionally used in experimental biology.

We think that this work could act as a good starting point for building scalable computer vision algorithms for motion processing that are rooted in biology.

This study has shown that a V1-MT feedforward model, together with the improvements described in this paper, can be successfully used to compute optic flow also in modern datasets. Moreover, it has opened up several interesting sub problems, which could be of relevance to biologists as well, for example what could be afferent pooling strategy of MT when there are multiple surfaces or occlusion boundaries within the MT receptive field? Can a better dense optical flow map be recovered by considering different multi-scale strategies? These questions are currently under consideration.

Acknowledgments

KM and PK acknowledge funding from the EC IP project FP7-ICT-2011-8 no. 318723 (MathEMACS)

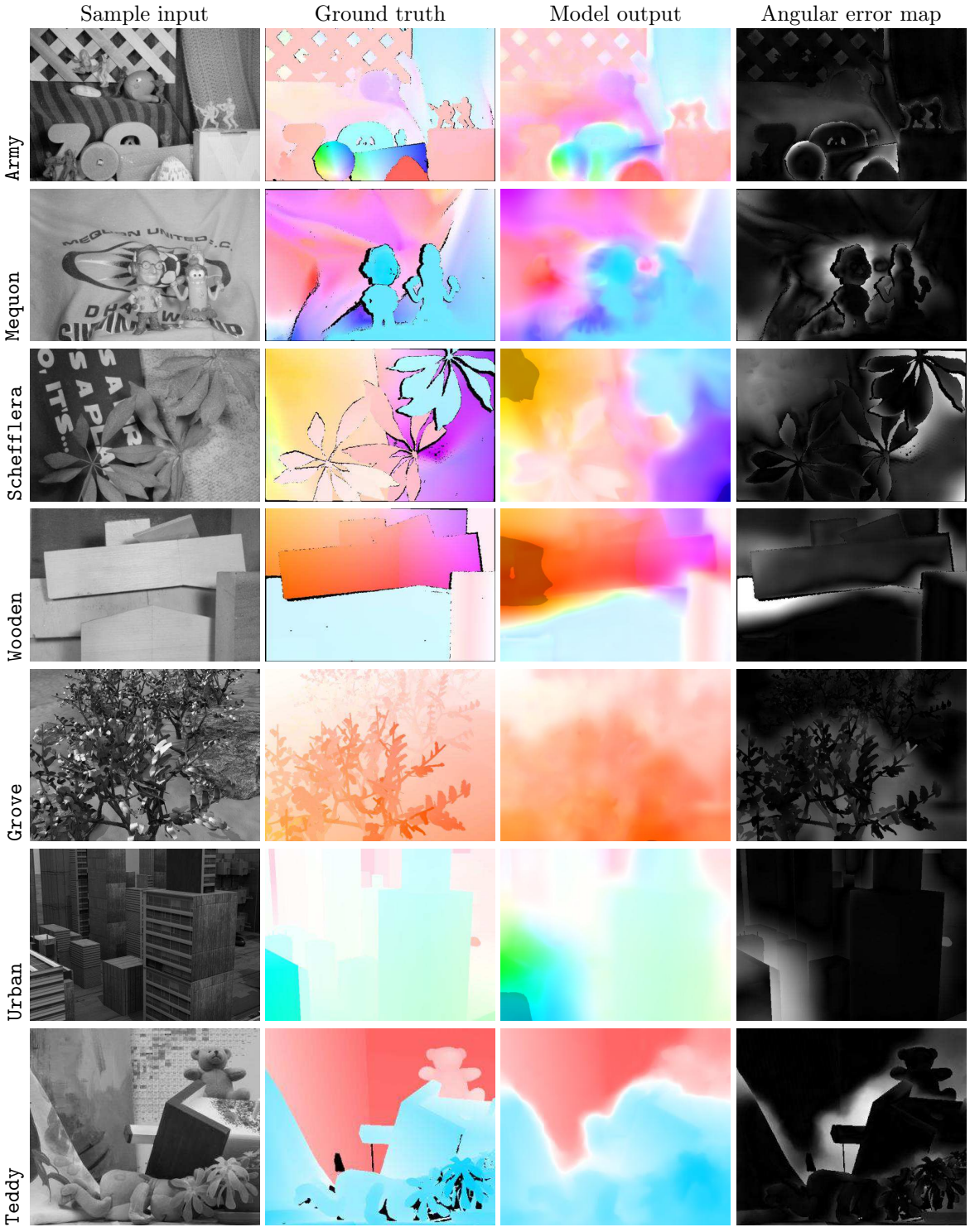


Figure 11: Sample results on Middlebury test set

Horn & Schunck [3]	86.6	8.01 89	19.9 90	8.38 90	9.13 86	23.2 93	7.71 84	14.2 86	25.9 89	14.6 90	12.4 92	30.6 91	11.3 91	4.64 90	5.64 90	4.60 87	8.21 90	24.4 95	8.45 91	4.01 80	5.41 85	1.95 24	9.16 92	17.5 93	8.86 89
SILK [79]	87.4	9.34 91	20.4 91	10.5 94	10.4 89	21.9 90	10.3 88	16.0 91	27.5 93	14.5 89	10.3 88	29.0 90	8.54 88	4.81 93	5.65 91	5.56 92	9.41 94	25.4 97	8.74 92	2.79 83	3.68 38	4.62 84	10.9 95	17.8 95	12.3 95
TI-DOFE [24]	89.9	13.4 99	23.2 97	16.5 100	16.5 96	24.1 95	18.2 98	20.2 100	31.1 100	20.6 99	19.9 99	32.9 96	20.8 100	4.89 94	5.90 94	5.54 91	8.04 89	23.9 94	8.81 93	2.97 60	4.34 62	1.88 20	10.9 95	17.7 94	11.9 93
HCIC-L [100]	92.5	15.7 101	22.0 94	10.1 93	31.5 102	26.6 99	41.0 102	14.8 90	23.1 83	16.8 94	18.4 98	34.4 98	18.2 99	5.94 98	6.35 98	6.35 95	10.6 97	19.2 81	11.4 97	18.7 102	17.8 102	19.2 101	4.93 72	8.34 82	5.16 72
V1-MT-FF [102]	92.8	13.2 97	22.6 95	9.37 92	13.6 93	31.2 102	10.5 90	17.0 94	28.1 94	11.0 81	16.4 97	32.8 95	15.2 93	5.91 97	6.07 97	7.03 97	21.3 101	36.9 101	23.4 101	3.73 77	5.25 84	2.86 55	12.3 97	18.8 98	14.8 96
SLK [47]	93.1	11.6 93	26.0 100	14.6 99	15.3 95	25.0 97	17.5 96	17.8 97	30.1 98	18.1 97	25.4 102	33.6 97	28.0 102	5.25 95	5.90 94	7.03 97	10.3 96	27.4 99	10.6 96	2.89 57	4.47 67	2.94 63	14.9 99	20.7 99	18.8 98
Adaptive flow [45]	94.8	13.2 97	20.8 92	14.0 98	17.1 98	22.0 91	17.9 97	18.1 98	27.1 91	22.8 101	11.8 89	31.1 92	10.5 89	6.35 100	7.13 100	6.25 94	9.87 95	21.8 91	9.44 94	12.6 101	11.4 101	20.0 102	7.75 91	13.6 87	7.73 87
PGAM+LK [55]	95.8	11.8 96	25.6 98	13.9 97	14.8 94	24.4 96	16.7 95	13.2 85	24.0 86	15.0 91	16.2 96	41.2 102	15.3 96	5.40 96	5.45 89	8.10 99	12.3 99	26.5 98	12.1 98	7.42 99	8.24 100	7.87 98	13.2 98	18.3 97	19.4 99
Periodicity [78]	96.7	11.2 94	27.0 101	7.46 87	16.6 97	29.8 101	18.2 98	25.3 102	31.2 102	24.9 102	12.7 93	35.7 100	11.1 90	31.7 102	41.4 102	25.1 102	23.8 102	41.5 102	23.8 102	2.92 58	5.62 86	6.90 93	18.6 101	33.1 102	22.3 100
FOLKI [16]	97.2	10.5 93	25.6 98	11.9 96	20.9 100	26.2 98	26.1 100	17.6 95	31.1 100	16.5 93	15.4 95	32.6 94	16.0 97	6.16 99	6.53 99	9.07 100	12.2 98	29.7 100	13.0 99	4.67 92	5.83 91	9.41 96	18.2 100	22.8 100	25.1 101
Pyramid LK [2]	99.3	13.9 100	20.9 93	21.4 102	24.1 101	23.1 92	30.2 101	20.9 101	29.5 97	21.9 100	22.2 101	34.6 99	25.0 101	18.7 101	23.1 101	20.2 101	21.2 100	24.5 96	21.0 100	6.41 97	7.02 97	10.8 98	25.6 102	31.5 101	34.5 102

Figure 12: A screenshot of the results for the Middlebury test set around our submitted model (V1-MT-FF).

References

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

[2] J. A. Perrone, A. Thiele, Speed skills: measuring the visual speed analyzing properties of primate MT neurons, Nature Neuroscience 4 (5) (2001) 526–532.

[3] C. C. Pack, R. T. Born, Temporal dynamics of a neural solution to the aperture problem in visual area MT of macaque brain, Nature 409 (2001) 1040–1042.

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

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

[6] P. Bayerl, H. Neumann, A fast biologically inspired algorithm for recurrent motion estimation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 29 (2) (2007) 246–260. doi:10.1109/TPAMI.2007.24.

[7] E. Tlapale, G. S. Masson, P. Kornprobst, Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism, Vision Research 50 (17) (2010) 1676–1692. doi:10.1016/j.visres.2010.05.022.
URL http://dx.doi.org/10.1016/j.visres.2010.05.022

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

[9] J. Bouecke, E. Tlapale, P. Kornprobst, 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, special issue on Biologically inspired signal processing: Analysis, algorithms, and applications. doi:10.1155/2011/781561.
URL http://www.hindawi.com/journals/asp/2011/781561.html

[10] D. Heeger, Optical flow using spatiotemporal filters, The International Journal of Computer Vision 1 (4) (1988) 279–302.

[11] S. Nowlan, T. Sejnowski, Filter selection model for motion segmentation and velocity integration, J. Opt. Soc. Am. A 11 (12) (1994) 3177–3199.

[12] E. Simoncelli, D. Heeger, A model of neuronal responses in visual area MT, Vision Research 38 (1998) 743–761.

[13] L. C. Sincich, J. C. Horton, The circuitry of v1 and v2: Integration of color, form, and motion, Annual Review of Neuroscience 28 (1) (2005) 303–326, pMID: 16022598. doi:10.1146/annurev.neuro.28.061604.135731.
URL http://dx.doi.org/10.1146/annurev.neuro.28.061604.135731

[14] M. J. Rasch, M. Chen, S. Wu, H. D. Lu, A. W. Roe, Quantitative inference of population response properties across eccentricity from motion-induced maps in macaque v1, Journal of Neurophysiology 109 (5) (2013) 1233–1249. doi:10.1152/jn.00673.2012.

[15] N. Rust, V. Mante, E. Simoncelli, J. Movshon, How MT cells analyze the motion of visual patterns, Nature Neuroscience 9 (2006) 1421–1431.

[16] J. Perrone, R. Krauzlis, Spatial integration by mt pattern neurons: a closer look at pattern-to-component effects and the role of speed tuning, Journal of Vision 8 (9) (2008) 1–14.

[17] D. Bradley, M. Goyal, Velocity computation in the primate visual system, Nature Reviews Neuroscience 9 (9) (2008) 686–695.

[18] C. Pack, R. Born, 2.11 - cortical mechanisms for the integration of visual motion, in: R. H. Masland, T. D. Albright, T. D. Albright, R. H. Masland, P. Dallos, D. Oertel, S. Firestein, G. K. Beauchamp, M. C. Bushnell, A. I. Basbaum, J. H. Kaas, E. P. Gardner (Eds.), The Senses: A Comprehensive Reference, Academic Press, New York, 2008, pp. 189 – 218. doi:http://dx.doi.org/10.1016/B978-012370880-9.00309-1.
URL http://www.sciencedirect.com/science/article/pii/B9780123708809003091

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

[20] J. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America A 2 (1985) 1160–1169.

[21] E. Adelson, J. Bergen, Spatiotemporal energy models for the perception of motion, Journal of the Optical Society of America 2 (1985) 284–321.

[22] N. Grzywacz, A. Yuille, A model for the estimate of local image velocity by cells in the visual cortex, Proceeding of the Royal Society of London B 239 (1990) 129–161.

[23] G. R. Stoner, T. D. Albright, V. S. Ramachandran, Transparency and coherence in human motion perception., Nature 344 (6262) (1990) 153–155.

[24] A. Noest, A. Van Den Berg, The role of early mechanisms in motion transparency and coherence, Spatial Vision 7 (2) (1993) 125–147.

[25] B. C. Skottun, Neuronal responses to plaids, Vision Research 39 (12) (1999) 2151 – 2156.

[26] G. C. Deangelis, I. Ohzawa, R. D. Freeman, Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation, Journal of Neurophysiology 69 (4) (1993) 1118–1135.

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

[28] J. H. Maunsell, 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) (1983) 1127–1147.

[29] K. R. Rad, 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, K. Q. Weinberger (Eds.), NIPS, 2011, pp. 846–854.

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

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

[32] S. Paris, P. Kornprobst, J. Tumblin, F. Durand, Bilateral filtering: Theory and applications, Foundations and Trends in Computer Graphics and Vision 4 (1). doi:10.1561/0600000020. URL http://dx.doi.org/10.1561/0600000020

[33] S. Ringbauer, S. Tschechne, H. Neumann, Mechanisms of adaptative spatial integration in a neural model of cortical motion processing, in: Proc. 10th International Conference on Adaptative and Natural Computing Algorithms (ICANNGA), LNCS, Springer, 2011.

[34] C. A. J.R. Bergen, E.H. Adelson, P. Burt, J. Ogden, Pyramid methods in image processing, RCA Engineer 29 (1984) 33–41.

[35] E. P. Simoncelli, Course-to-fine estimation of visual motion, in: IEEE Eighth Workshop on Image and Multidimensional Signal Processing, 1993.

[36] G. C. DeAngelis, I. Ohzawa, R. D. Freeman, Receptive-field dynamics in the central visual pathways, Trends in Neurosciences 18 (10) (1995) 451 – 458.

[37] D. A. Clausi, M. E. Jernigan, Designing Gabor filters for optimal texture separability, Pattern Recognition 33 (11) (2000) 1835 – 1849.

[38] T. D. Albright, R. Desimone, Local precision of visuotopic organization in the middle temporal area (MT) of the macaque, Experimental Brain Research 65 (3) (1987) 582–592.

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

Contents

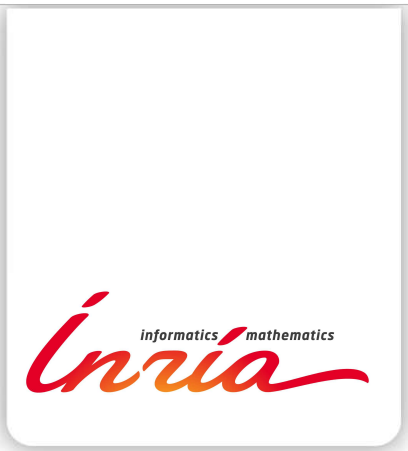
1 Introduction **3**

2 Model description **4**
2.1 Context: Biological visual motion estimation 4
2.2 Model overview 4
2.2.1 Step 1 : V1 (*Motion energy estimation and normalization*) 4
2.2.2 Step 2: MT pattern cells response 6
2.2.3 Step 3: Velocity estimation 6
2.3 Extended approaches 7
2.3.1 Bilateral filter solution (BF) 7
2.3.2 Trilateral filter solution (TF) 8

3 Algorithmic details: How to make the approach applicable to real sequences **8**
3.1 Multiscale approach 8
3.2 Boundary conditions 9
3.3 Unreliable regions 10

4 Results **10**
4.1 Parameters settings 10
4.2 Analysis of proposed approaches 11
4.3 Performance evaluation on Middlebury dataset 16

5 Conclusion **16**



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