# Benchmarking biologically inspired spatio-temporal filter based optical flow estimation on modern datasets

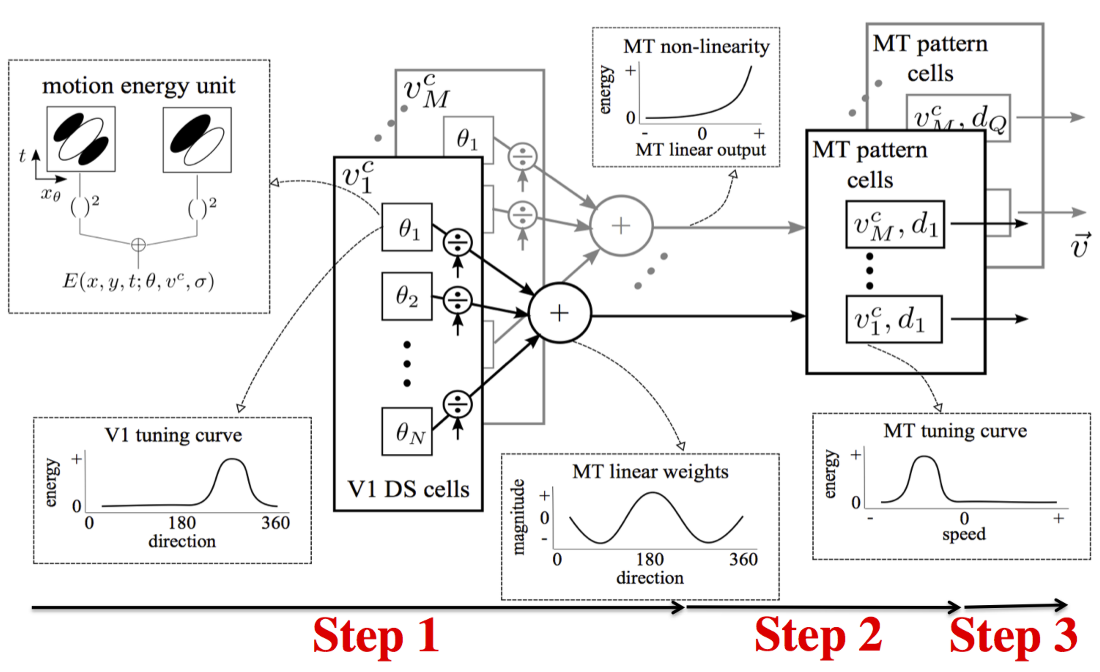
**Fabio Solari(1), Manuela Chessa(1), N V Kartheek Medathati(2) - Pierre Kornprobst(2)**
(1) University of Genoa, Italy – (2) Inria, Team Neuromathcomp, France

## Context and motivation

- Visual motion information is useful for a multiplicity of tasks.
- Analysis of it is of great interest both in machine vision for developing applications and biological vision as is it subserves several functional needs.
- Owing to the application potential, the problem of visual motion estimation has also received a lot of attention from the computer vision and robotics researchers, see [3, 12] reviews.
- One of the best practice in the computer vision community is benchmarking of various algorithms on publicly available datasets for e.g. Middlebury [2], MPI-Sientel [5] which played a crucial role in rapid development of algorithms with increasing accuracy.
- Neural mechanisms underlying motion analysis have been studied extensively in the past 2 decades, see [11, 7, 4] for reviews.
- Mathematical models describing the neural mechanisms [1, 10, 13] have received little attention in terms of their performance on computer vision datasets,
  - How do these models perform on the modern benchmarking datasets?
  - Are they scalable?
- We begin to look into this question by examining feedforward spatio-temporal filter based models proposed by Heeger [10].

## Estimation of Motion Energy and Velocity

▶ Overview of the model for velocity estimation




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

$$\boldsymbol{E^{V1}(x,y,t;\theta,v^c,\sigma)} = \frac{E(x,y,t;\theta,v^c,\sigma)}{\sum_{\theta_i=\theta_1}^{\theta_N} E(x,y,t;\theta_i,v^c,\sigma)+\epsilon} \quad (1)$$

where,

$$\begin{aligned} E(x,y,t;\theta,v^c,\sigma) &= R_e(x,y,t;\theta,v^c,\sigma)^2 + R_o(x,y,t;\theta,v^c,\sigma)^2 \\ R_{e/o} &= g_{e/o} * I(x,y,t) \\ g_e(x,y,t) &= h_e(x,y)p_e(t) - h_o(x,y)p_o(t) \\ g_o(x,y,t) &= h_o(x,y)p_e(t) + h_e(x,y)p_o(t) \\ h(x,y) &= Be^{\left(\frac{-(x^2+y^2)}{2\sigma^2}\right)} e^{j2\pi f_{x0}x+f_{y0}y} \\ p(t) &= e^{\left(-\frac{t}{\tau}\right)} e^{j2\pi(f_{t0}t)} \end{aligned}$$

▶ Step 2 : MT (*Motion energy pooling and non-linearity*)

$$\boldsymbol{E^{MT}(x,y,t;v^c,d,\sigma)} = F\left(\sum_{\theta_i=\theta_1}^{\theta_N} w_d(\theta_i) G * E^{V1}(x,y,t;\theta_i,v^c,\sigma)\right) \quad (2)$$

where,

$$w_d(\theta_i) = cos(d-\theta_i)$$

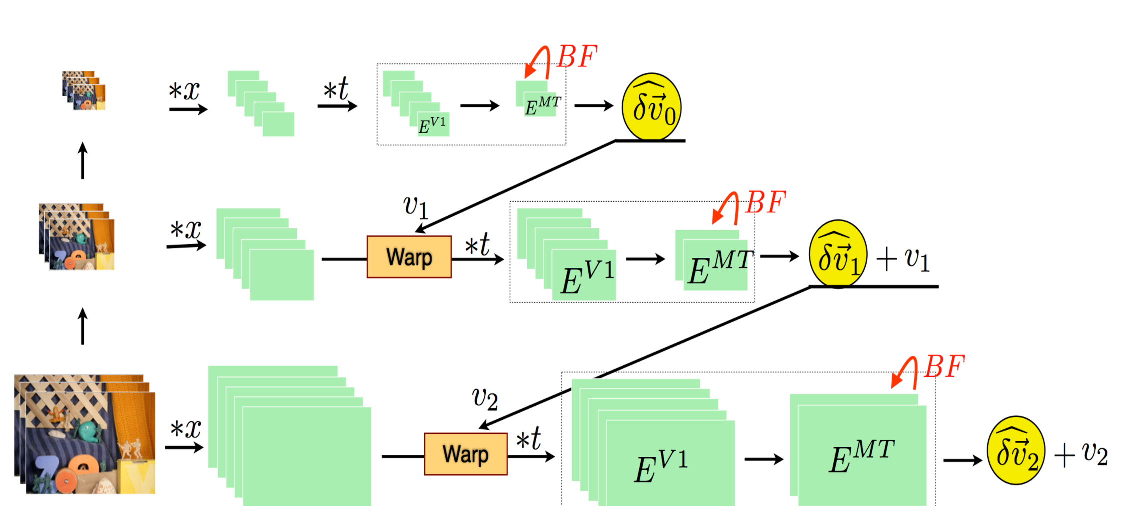
▶ Step 3 : Velocity estimation

$$\boldsymbol{v_x} = \frac{\sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{MT}(x,y,t;v_i^c,0,\sigma)}{\sum_{v_i^c=v_1^c}^{v_M^c} E^{MT}(x,y,t;v_i^c,0,\sigma)} \qquad \boldsymbol{v_y} = \frac{\sum_{v_i^c=v_1^c}^{v_M^c} v_i^c E^{MT}(x,y,t;v_i^c,\pi/2,\sigma)}{\sum_{v_i^c=v_1^c}^{v_M^c} E^{MT}(x,y,t;v_i^c,\pi/2,\sigma)} \quad (3)$$

Compared to earlier formulation of Heeger, where velocity is computed using a least squared optimization procedure, it has been formulated into a linear-non-linear feedforward routine.

## Implementation Details

- Multi-scale method.
- Discontinuity enhancement using bilateral filtering at the level of MT energy.




## About simulations

▶ Parameters

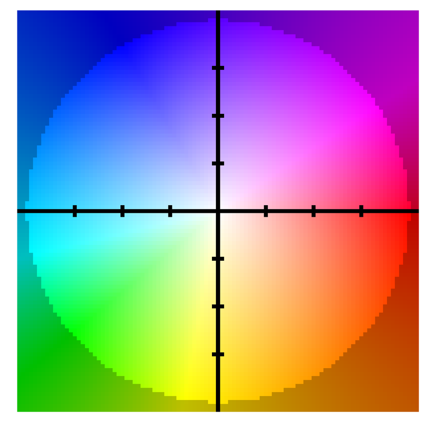
**V1**

- Spatial filter support: 11 pixels
- Temporal filter support: 7 frames
- $\tau = 2.5$ frames
- Orientation tuning: 8 in [0 to $\pi$]

**MT**

- Velocity tuning: [-0.9, -0.4, 0.0, 0.4, 0.9] ppf
- Orientation tuning: [0, $\pi/2$]

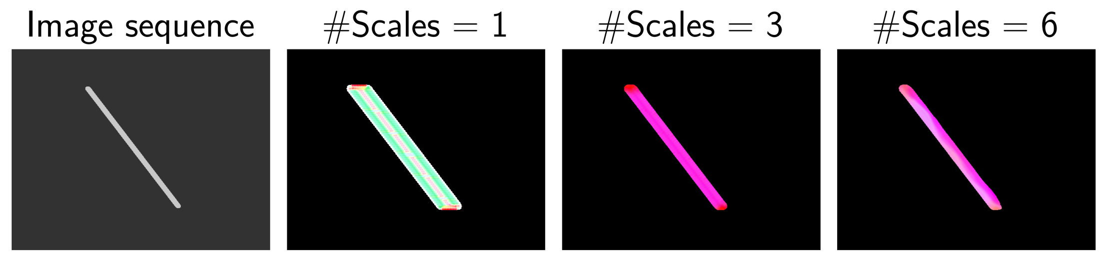
▶ Visualization: Velocity color code



## Tests on synthetic images

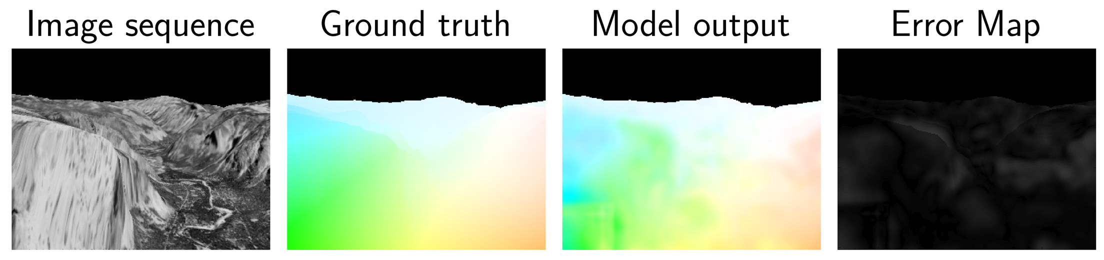
▶ Multi-scale helps solving motion integration problem

Image sequence | #Scales = 1 | #Scales = 3 | #Scales = 6



▶ Using Yosemite video sequence

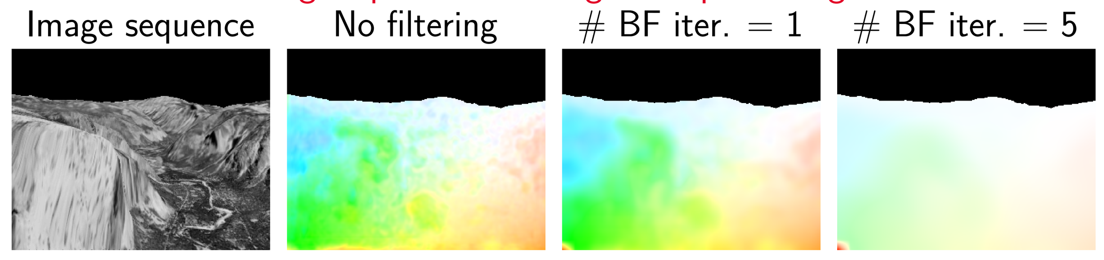
Image sequence | Ground truth | Model output | Error Map



AAE = 7.12,Std. AAE = 5.36, End-Point = 0.35, Std. End-Point = 0.37.

▶ Non linear filtering helps in smoothing while preserving discontinuities

Image sequence | No filtering | # BF iter. = 1 | # BF iter. = 5



AAE= 7.89, 5.81 AAE= 6.80, 5.26 AAE= 5.53, 5.31
EPE= 0.36, 0.32 EPE= 0.3, 0.28 EPE= 0.26, 0.43

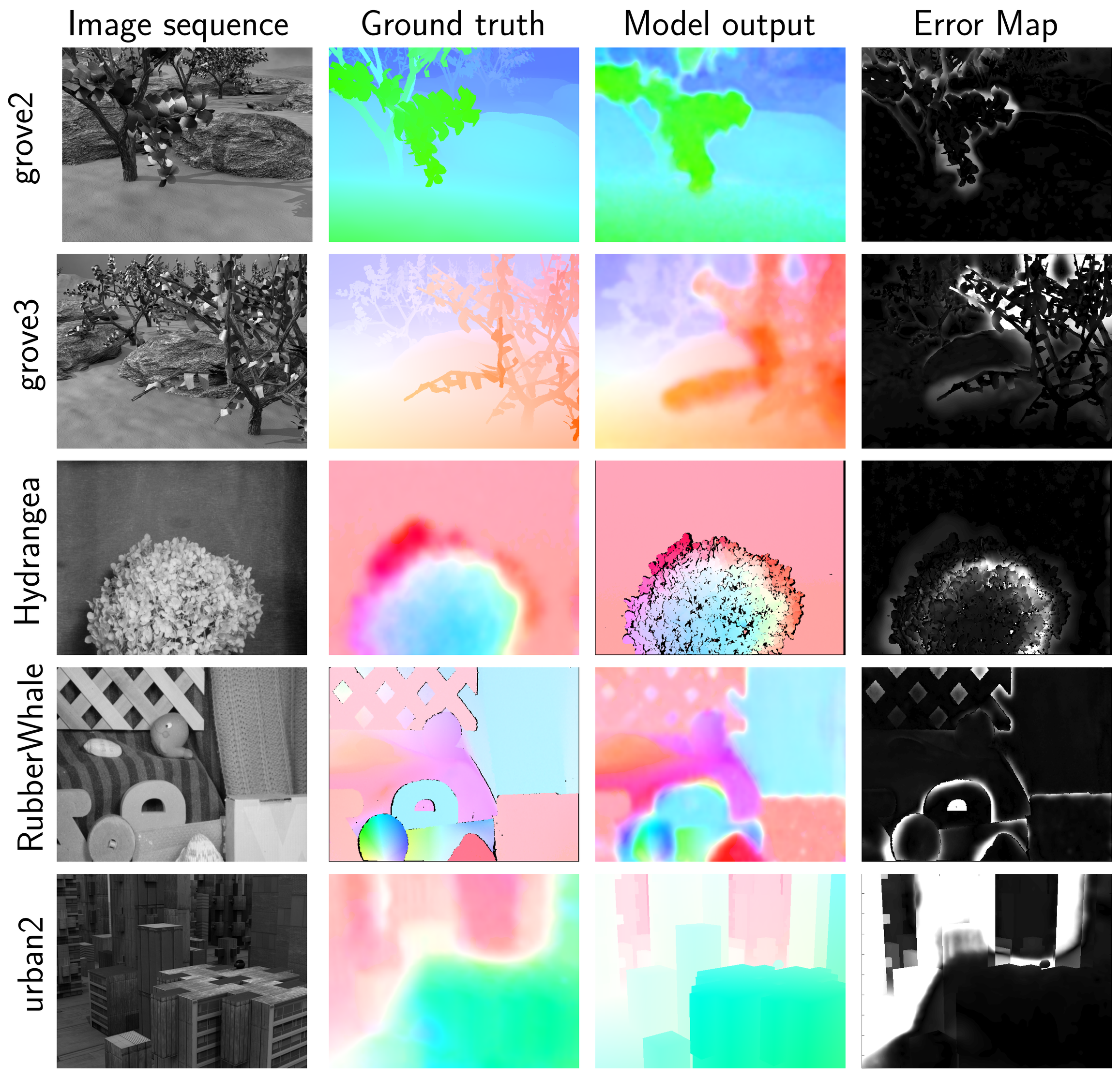
## Performance evaluation using Middlebury Dataset

Proposed in ICCV-2007, contains sequences with non-rigid motion where ground truth is obtained by tracking fluorescent texture and realistic synthetic sequences divided into training and testing sets.

▶ Results on training sequence

| Sequence | AAE | Std. AAE | End-Point | Std. End-Point |
|---|---|---|---|---|
| grove2 | 4.62 | 9.81 | 0.31 | 0.59 |
| grove3 | 10.37 | 19.38 | 1.30 | 1.88 |
| Hydrangea | 8.01 | 15.05 | 0.95 | 1.33 |
| RubberWhale | 11.13 | 17.24 | 0.37 | 0.53 |
| urban2 | 39.46 | 46.44 | 3.56 | 2.22 |
| urban3 | 20.60 | 43.88 | 2.01 | 3.30 |

▶ Illustration: Results on training sequence




▶ Analysis

- The model does not have implicit mechanism to deal with blank-wall problem but is helped by multi-scale approach.
- Results are better regions with coarse texture.
- Changes in velocity with-in support of temporal filter leads to erroneous estimates.

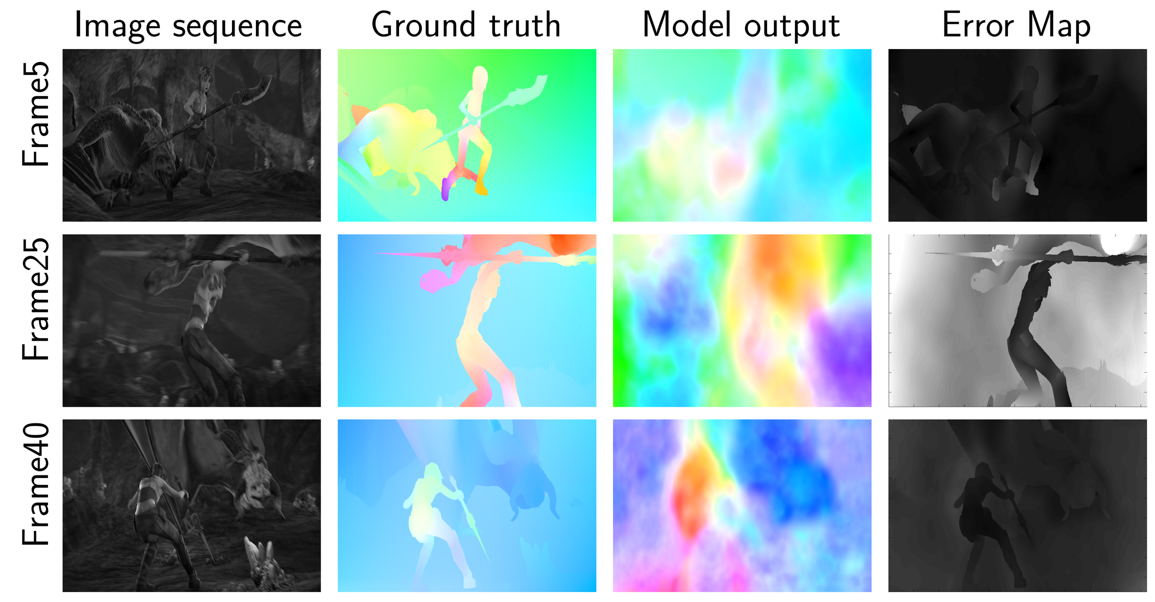
## MPI-Sientel Dataset

Proposed in ECCV-2012, contains longer synthetic sequences that match statistics of natural images and includes challenging effects such as blur, transparency and specular reflections.

▶ Results on cave2 training sequence

| Sequence | AAE | Std. AAE | End-Point | Std. End-Point |
|---|---|---|---|---|
| Frame 5 | 24.51 | 26.95 | 10.60 | 5.55 |
| Frame 25 | 60.21 | 42.75 | 61.23 | 17.75 |
| Frame 40 | 53.16 | 42.13 | 20.06 | 4.97 |

▶ Illustration: Results on training sequence




## Discussion

- Spatio-temporal energy features have been very recently demonstrated to be effective in high level vision applications such as Dynamic video segmentation[8], human action recognition[9], region tracking [6].
- They have also been successful in modeling works in neurophysiology and psychophysics, so they truly lie at the interface of both understanding biology and developing scalable computer vision applications.
- Even though early models such as [10] have been proposed for dense optical flow estimation, very little is known in terms of their performance on modern computer vision datasets.
- In this we have benchmarked one model as a starting point.

## Future work

- Public domain datasets and code sharing practices have proved to be very effective in advancing computer vision algorithms.
- Developing code repository for biologically inspired models and benchmarking them on computer vision datasets.
- Handling boundary conditions with appropriate extrapolation methods.
- Handling low energy regions using diffusion mechanism.
- Using machine learning techniques for read-out to obtain locally context aware velocity estimation.

## References

[1] E.H. Adelson, et. al. Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America* , 1985.

[2] S. Baker, et. al. A database and evaluation methodology for optical flow. In *International Conference on Computer Vision*, 2007.

[3] J.L. Barron, et. al. Performance of optical flow techniques. *The International Journal of Computer Vision*, 1994.

[4] D.C. Bradley et. al. Velocity computation in the primate visual system. *Nature Reviews Neuroscience*, 2008.

[5] Daniel J. et. al. A naturalistic open source movie for optical flow evaluation. In *Proceedings of European Conference on Computer Vision*, 2012.

[6] K.J. Cannons et. al. The applicability of spatiotemporal oriented energy features to region tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2014.

[7] C.W.G Clifford, et. al. Fundamental mechanisms of visual motion detection: models, cells and functions. *Progress in Neurobiology*, 2002.

[8] T. Damien, et. al. Segmentation of dynamic scenes with distributions of spatiotemporally oriented energies. In *Proceedings of BMVC*, 2014.

[9] M. J. Escobar, et. al. Action recognition via bio-inspired features: The richness of center-surround interaction. *Computer Vision and Image Understanding*, 2012.

[10] D.J. Heeger. Optical flow using spatiotemporal filters. *The International Journal of Computer Vision*, 1988.

[11] K. Nakayama. Biological image motion processing: A review. *Vision Research*, 1984.

[12] D. Sun, et. al. Secrets of optical flow estimation and their principles. *Proceedings of CVPR*, 2010.

[13] E. Tlapale, et. al. A neural field model for motion estimation. In Springer Verlag, editor, *Mathematical Image Processing*, 2011.

N V Kartheek Medathati
kartheek.medathati@inria.fr

This work was partially supported by the
EC IP project FP7-ICT-2011-8 no. 318723 (MatheMACS)