Motion Field Diffusion Model Controlled by Form Cues

Émilien Tlapale[†], Guillaume S. Masson[‡], Thierry Viéville[†] and Pierre Kornprobst[†]

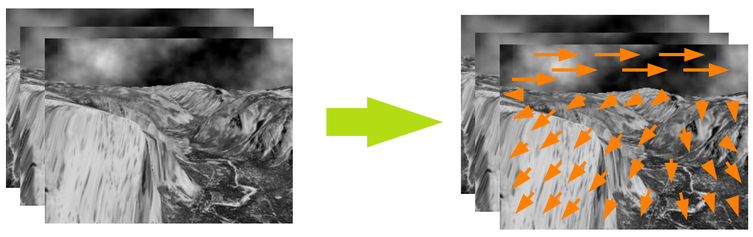
[†] Projet Odyssée, INRIA Sophia-Antipolis, ENS, ENPC
[‡] Équipe DyVA, INCM CNRS

Contact: Emilien.Tlapale@sophia.inria.fr
http://emilien.tlapale.com

We present a model of motion integration and segmentation inspired from the primate visual system. Local motion estimation is solved through a recurrent V1-MT process. However, the feedback diffusion is controlled via a second recurrent network (V1-V2 like) that estimates local form information at a different spatial scale.

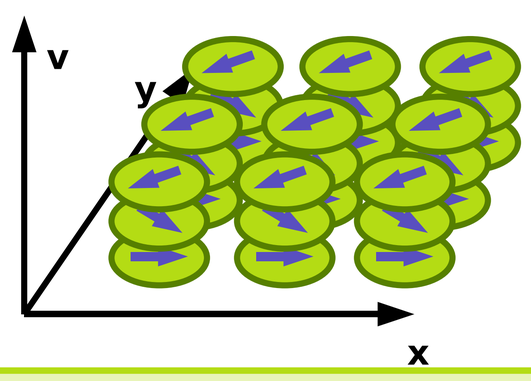
Optical Flow

- Given a frame
- sequence, we want
- to estimate the
- **pattern** motion
- at each point

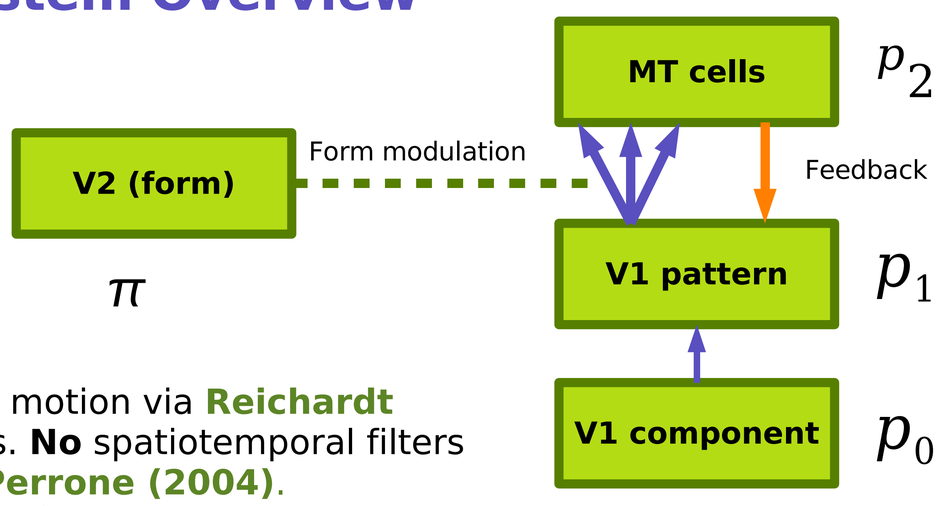


- V1 and MT neurons are **velocity**
- **tuned**.

- As the probabilistic model of
- **Weiss (1998)** or **Bayerl &**
- **Neumann (2004)**'s
- dynamical system we want to
- keep this property



System Overview



Local motion via **Reichardt** filters. **No** spatiotemporal filters like **Perrone (2004)**.

Low velocities prior

Feedback from MT
Used for center-surround
(Angelucci & Bullier, 2003)

$$p_1 = p_0 \hbar(v)(1 + c_f p_2)$$

$$p_2 = \int_\xi G_{\sigma(\kappa(\xi))}(x-\xi)\,\pi(\xi, \Theta(\overrightarrow{\xi x}))\,p_1(\xi, v)$$

MT receptive fields are larger than V1

Lower contrast implies larger surrounds **(Tadin et al., 2003, Pack et al., 2005)**

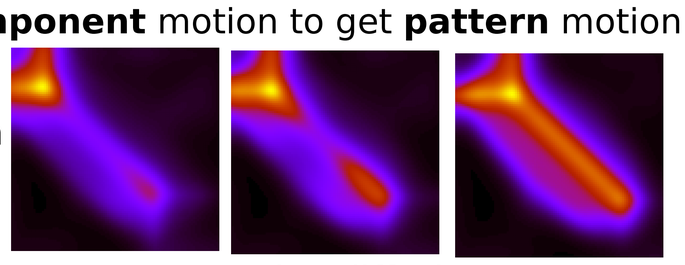
Form modulation

MT's main input is V1 local motion detectors

Motion Integration

Integration combines **component** motion to get **pattern** motion

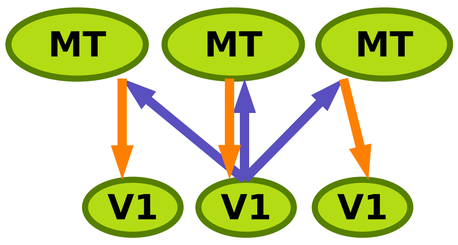
- **Pack et al. (2003)** suggest motion perception depends on terminators



Motion diffusion MT

- **Bayerl & Neumann (2004)** propose a *recurrent system for* motion diffusion

- In our model, **center-surround** interaction in V1 is done via feedback as described by **Angelucci & Bullier (2003)** as are MT horizontal connections



Form Modulation

- **Huang et al. (2007)** provide physiological data on form modulation for motion processing
- **Weiss (1998)** and **Bayerl & Neumann (2004)** models ignore junctions (T and X or tricolored) to avoid implicit ones
- Our model using a direction of figure (different from **Sajda & Baek, 2004**) creates asymetric center surround depending on the stimuli

Our **π** encodes the **direction of figure**

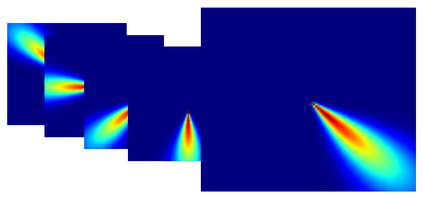
$$\pi(x, d) = \int_\xi G_{\sigma(\kappa(x))}(\xi - x, d - \Theta(\overrightarrow{x\xi}), S(\xi, x))$$

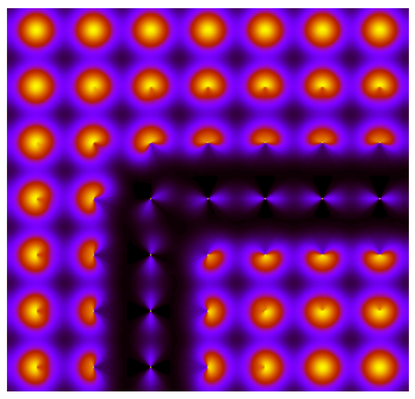
Contrast dependance

Spatial & angular distances

Similarity

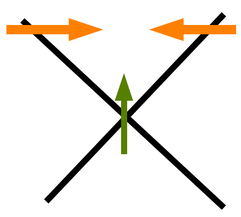
Directional Gaussians
= Angular + Spatial distance

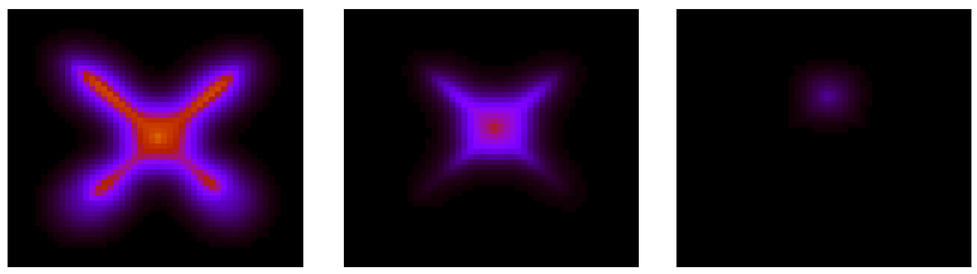




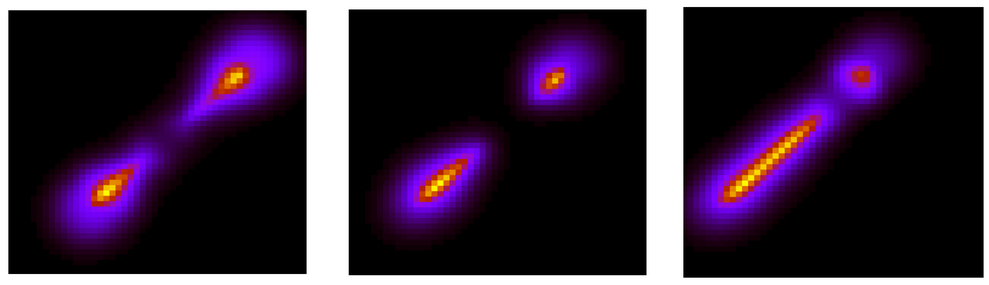
Results

- Crossing bars: the intersection
- is an **extrinsic junction**



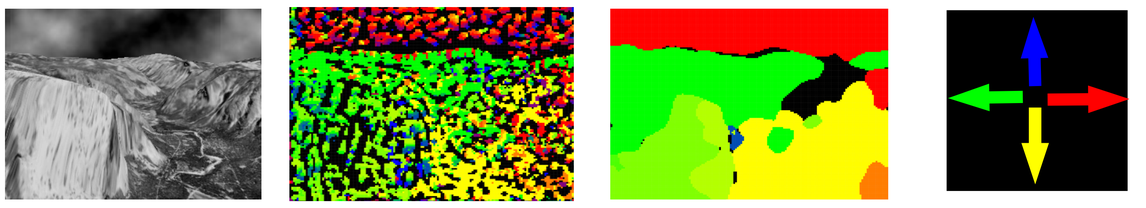


Upward motion MT cells



Leftward motion MT cells

- Unlike **Weiss (1998)** or **Bayerl & Neumann (2004)** we do not build a junction detector but use the direction of figure to manage extrinsic junctions



Extracted V1 and MT higher motion responses on Yosemite frames

Acknowledgments: This work was partially supported by the EC IP project FP6-015879, FACETS