Mechanisms of Adaptive Spatial Integration in a Neural Model of Cortical Motion Processing

Stefan Ringbauer, Stephan Tschechne, and Heiko Neumann

Ulm University
Faculty of Engineering and Computer Science
Institute for Neural Information Processing
89069 Ulm, Germany
{stefan.ringbauer,stephan.tschechne,heiko.neumann}@uni-ulm.de

Abstract. In visual cortex information is processed along a cascade of neural mechanisms that pool activations from the surround with spatially increasing receptive fields. Watching a scenery of multiple moving objects leads to object boundaries on the retina defined by discontinuities in feature domains such as luminance or velocities. Spatial integration across the boundaries mixes distinct sources of input signals and leads to unreliable measurements. Previous work [6] proposed a luminance-gated motion integration mechanism, which does not account for the presence of discontinuities in other feature domains. Here, we propose a biologically inspired model that utilizes the low and intermediate stages of cortical motion processing, namely V1, MT and MSTl, to detect motion by locally adapting spatial integration fields depending on motion contrast. This mechanism generalizes the concept of bilateral filtering proposed for anisotropic smoothing in image restoration in computer vision.

Keywords: Motion Estimation, Neural Modeling, Motion Integration, Diffusion.

1 Introduction

Optic flow is perceived whenever observer motion occurs relative to the surrounding environment. The resulting projected movements are caused by object motion in the scene or egomotion, or combinations of both. The primate brain has built an impressive ability to derive rich information from such data, making navigation in and interaction with the environment reliable and accurate. As many applications could profit from flow-based information, motion estimation has been subject to intense research in the past decades (e.g. [5], [3], [1]). Biological models are a promising way towards reliable motion estimation and have already demonstrated results comparable with engineering approaches against changes in complex scenes. In accordance with the structure of the the visual system biologically inspired models consist of interconnected areas with specialized functionality. Recurrent connections between such areas help improving and stabilizing the derived motion signal.

A. Dobnikar, U. Lotrič, and B. Šter (Eds.): ICANNGA 2011, Part I, LNCS 6593, pp. 110–119, 2011.

The hierarchical integration over increasingly larger neighborhoods causes the problem that information is mixed up at boundaries where different input sources might contribute to the integration field. In such cases the feedforward integration leads to an erroneous measure of surface motion. Some approaches have already suggested improvements of such integration steps, e.g. by using luminance contrast to modulate integration [6]. However, this only works if luminance values of objects and background differ sufficiently, an assumption that only holds for rather limited number of cases. In this paper we propose an adaptation of the integration mechanism to render it sensitive to motion discontinuities. Shapes of receptive fields for motion integration are adapted in an anisotropic fashion to integrate coherently moving regions but to prevent integration across motion discontinuities. We adapted a model of motion estimation [4] which consists of a hierarchical architecture of the dorsal pathway containing the primary visual area V1, medial temporal area MT, as well as medial superior temporal area MST. Our method alters the motion integration step between area V1 and MT. This process includes the estimation of an anisotropic filter shape by utilizing motion contrast information from area MST to a diffusion-like estimation of the corresponding integration region of cells in MT. With this extension, the integration regions adapt to the outline of moving objects, independent of luminance or other features. This improves quality of motion estimation at object borders and helps to disambiguate further processing.

2 Neural Model

The model proposed is based on [4] which is a recurrent neural model covering the main stages of the motion processing (dorsal) pathway of the primate brain. The model consists of areas V1, MT and the ventral part of MST, namely MSTv. Model area V1 is fed with image sequences and incorporates correlation-based initial motion estimation where area MT considers homogeneous motion integration. It also supports the initial estimates at V1 level by feeding back contextual information obtained from integrating motion information over large spatial neighborhoods. This results in motion information that is represented in form of population codes that is finally interpreted to receive an optical flow field. We extend this model by incorporating the lateral part of area MST, namely MSTl, with cells sensitive to motion discontinuities. This motion contrast information is needed for the actual extension to incorporate a motion contrast dependent motion integration mechanism from area V1 to area MT.

2.1 Method for Motion Discontinuity Detection

MT motion activities are further processed by contrast sensitive cells in model area MSTl [8]. The cells have approximately the same size as MT cells and are selective for both direction and speed changes of the input motion responses. This is done by looking for different motions in the spatial neighborhood of a cell. The receptive fields of these motion contrast cells are composed of a

small center and a larger surround area using Gaussian kernels G_σ as weighting function to yield $center = act^{MT} * G_c$ and $surround = act^{MT} * G_s$, respectively (where $*$ denotes the convolution operation). For each spatial position motion contrast is computed by a divisive inhibition

$$act^{MSTl}{}_{[\theta|s]}(x) = center_{[\theta|s]}(x)/(\eta + \beta \cdot surround_{[\theta|s]}(x)) \tag{1}$$

where θ is the motion direction co-domain with cardinality k, s the speed co-domain with cardinality l, η is a scaling constant, and β is a constant for surround modulation. Locations with high activation in model MSTl represent motion discontinuities that are transitions between different motions. In order to detect such contrast locations irrespective of the composite input velocity, we sum the contrast signals from the direction and speed co domains, namely

$$act^{MSTl}(x) = \sum_{i=1}^{k} \frac{act^{MSTl}{}_{\theta}(x)}{k} \cdot \sum_{j=1}^{l} \frac{act^{MSTl}{}_{s}(x)}{l} \tag{2}$$

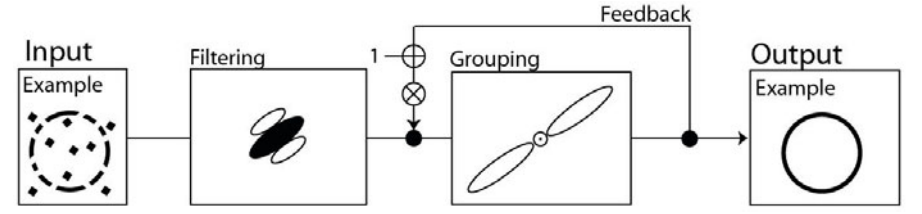
This information will be used to limit the integration areas to avoid a mix-up of integrating different motions.

2.2 Model of Contextual Enhancement

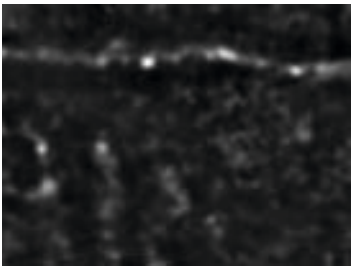
The estimation of motion discontinuities is likely to be noisy and contextual boundaries might be fragmented. The unenhanced signal thus does not clearly separate regions of different motion, as desired. We propose a step of modulatory enhancement that reduces the noise and mends discontinuities that belong together. This process is based on the linking principle first proposed by [13] which was used and further extended by [11], [12], [14]. In a nutshell, a driving feedforward (FF) signal is enhanced by feedback (FB) signals following the scheme $act^{FF}(1 + \alpha\, act^{FB})$. Our proposed context enhancement consists of two interacting layers, which are designed following the layout of the early visual cortex. The first layer models cells that are sensitive for contrasts in a specific orientation and spatial frequencies. The second layer applies long-range connections to group like-oriented contrasts. Activity of the second layer is then used to enhance the signal from the first layer by applying a modulatory feedback signal. This enhances contextually meaningful contours that receive support from their surround while removing spurious activities. Please note that while the principle is adapted from the lower areas of visual cortex, it now processes signals originating from higher areas. For details, see Figure 1 that shows an illustration and examples of the process. In our experiments we used Gabor filters for the first stage, and a multiplicative combination of Gaussian-shaped cells for the second stage. They are used to build a population with 16 orientations each. The feedback iterations between both layers are repeated several times until gaps in contextual contours are closed sufficiently and until noise is removed.

2.3 Model of Diffusion Process

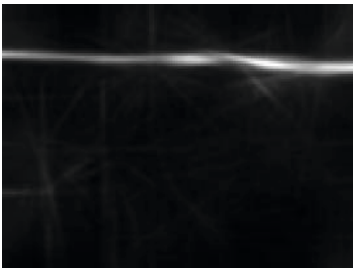
To infer the shape of the integration after the enhancement step, we incorporate a process of anisotropic diffusion which is carried out for every MT cell.



(a)



(b)



(c)

Fig. 1. Modulatory enhancement. (a) Model sketch of two interacting layers with examples (contrast orientation and grouping stage) (b) Example input (c) Result of the contextual enhancement. Undesired noise was cancelled while contextual clues have been increased.

During the diffusion process, activities of the modulatory enhancment steer the diffusion process. A diffusion process equilibrates concentration differences while preserving the summed amount of particles, mass or energy. In our case, *energy* or *mass* is the spatial weight of a location within the receptive field to contribute to the integration process. In formal terms, strength and direction of the resulting diffusion (*flux*) j depends on a diffusion tensor D and the gradient ∇u of the energy field. The property of energy conversation follows *Fick's Law* ([2], [7]) and leads to the expression of the diffusion equation

$$\partial_t u = -div\, j \quad \text{with} \quad j = -D \cdot \nabla u \tag{3}$$

Equation 3 realizes the heat conduction process if matrix D is the unit matrix with a homogeneous scaling constant. We utilize an anisotropic scheme in which the local direction is steered by the local orientation of the grouping responses. In this case, D is designed as follows. The predominant orientation θ is found with $\theta = \text{argmax}(act_{\theta_N}^{MSTl})$ for $N \epsilon \{1..orientations\}$. This is used to define basis vectors for the diffusion process $\mathbf{v1} = \begin{pmatrix} cos\,\theta \\ sin\,\theta \end{pmatrix}$ and $\mathbf{v2} = \begin{pmatrix} -sin\,\theta \\ cos\,\theta \end{pmatrix}$. The diffusion matrix D is then obtained after scaling using the eigenvalues

$$\lambda_1 = \begin{cases} 1 & act_\theta^{MSTl} = 0 \\ 1 - e^{-\frac{1}{act_\theta^{MSTl}}} & act_\theta^{MSTl} > 0 \end{cases} \text{ and } \lambda_2 = 1 \text{ in } D = \begin{pmatrix} \mathbf{v_1}\ \mathbf{v_2} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \mathbf{v_1^T} \\ \mathbf{v_2^T} \end{pmatrix}.$$

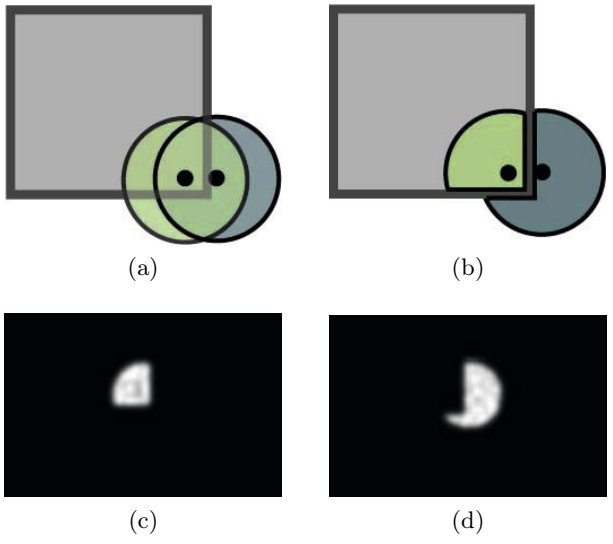


Fig. 2. Diffusion process. (a) Homogeneous integration would mix different features (illustrated by the box) when cells are close to discontinuities. (b) Anisotropic integration keeps integration regions separate. (c,d) Model simulations of an integration area located close to but on different sides of a discontinuity. In our proposal, contextually enhanced motion discontinuities steer the diffusion process.

2.4 Combination of Methods

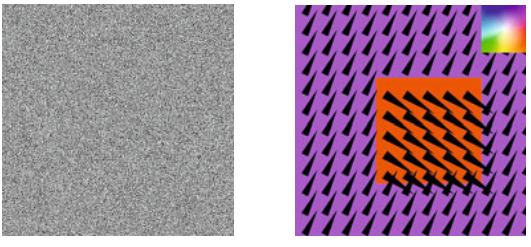
The previous sections introduced all mechanisms that where added to the underlying model proposed in [4]. These mechanisms are now combined. After the initial motion detection in model area V1 the information is integrated to the spatially coarser MT and passed on to MSTl where the motion is analyzed by its contrast sensitive cells (section 2). MSTl responses are then sharpened by the context enhancement mechanism (section 2.2). The resulting information about the motion discontinuities in MSTl is then fed back to model area MT that incorporates it into the motion contrast dependent integration (MCDI). The MCDI uses a diffusion process (section 2.3) for each receptive field that is bounded by the motion discontinuities calculated earlier. The receptive fields in turn selectively integrate only motion information from V1 that is covered by the region determined by the diffusion mechanism.

3 Results

The model proposed was probed with various selectively test input sequences testing its capabilities for the separation of different motions within the integration. This is shown using two different image sequences: First, an abstract artificial sequence containing moving noise patterns and second, the well known 'Yosemite with Clouds' sequence which is close to a real world example.

3.1 Moving Box

This image sequence is a composition of a rectangular patch moving to the lower right in front of a background that is moving to the upper right. Both, the patch and the background, are generated by random noise (equally distributed) luminance pattern. As a result the patch can only be recognized during component motion since no static form clues where available to detect region boundaries (Figure 3).

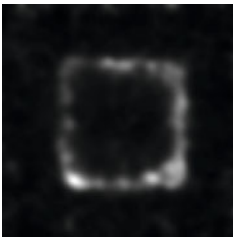


(a) original input frame (b) optical flow (ground truth)

Fig. 3. (a) The textured moving box and background. There are no static features to determine the borders of the moving box. (b) When in motion, the box show an outline and can be segmented from the background. Here the ground truth optical flow is displayed. The color encodes the motion direction as shown in the color wheel on the upper right.

Given this input sequence of surface motion our model generates strong responses in MSTl motion contrast cells at the transition between the box and the background (Figure 4a). There are also weak responses within the box and on the background due to noise in the motion estimation. After the stage of context enhancement the noise within the box and the background is reduced and the boundary of the box induced by motion contrast appears less fragmented (Figure 4b). These boundaries affect the integration process by limiting the integration region within a receptive field (RF) as can be seen in Figure 4c where the darkened areas show positions where the integration area is decreased through the motion contrast.

Comparing the estimated flow of the model using homogeneous (HI) and motion contrast dependent integration (MCDI) reveals the advantage of the selective motion integration. The HI produces rather smooth transitions at the boundaries of different motions whereas the MCDI leads to more differentiated and sharper boundaries such that regions of strong differences between estimated and ground truth flow in proximity of motion boundaries are greatly reduced (Figure 5). In case of the noise-textured object the projected error [10] is improved by an average of 18.7% with MDCI (measuring the first nine frames). An adaptation mechanism that utilizes luminance contrasts instead of motion would fail in this scenario. The sequence is composed of regions with the same



(a) motion contrast cell response

(b) context enhanced motion contrast

(c) percentage of RF integration area

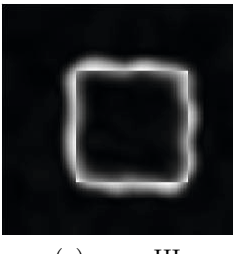
Fig. 4. (a) The response of the motion contrast sensitive cells in MSTl (lighter means higher activity) on the moving box sequence. The highest activities are found at the borders even though the boundary appears fragmented. (b) After context enhancement the boundaries of the boxes are intensified and the noise is highly reduced. (c) The percentage of the region of a RF integration is shown (reduced integration regions are darkened).

luminance statistics so that no form information helps to correlate with the moving patches.

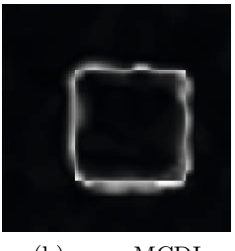
3.2 "Yosemite with Clouds" Sequence

The 'Yosemite with Clouds' image sequence simulates a flight over a virtual mountain scenery. One quarter of the image consists of sky with clouds moving to the right, the rest shows a canyon of the Yosemite National Park (hence the name) which passes under the observer as the camera moves forward and slightly rotates to the right.

When the estimated flow after homogeneous (HI) integration is compared to the one with motion contrast dependent integration (MCDI), the difference is minute. The difference becomes more striking if we focus on the population code instead of its interpretation. Let's zoom in to the upper left of the sequence,

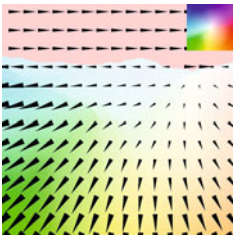


(a) error HI

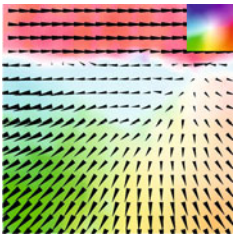


(b) error MCDI

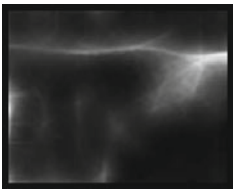
Fig. 5. (a) In quite a distance to the border of the box errors in estimated motion occur, since the homogeneous integration fused the background and foreground motion. (b) Erroneous estimations appear only very close to the motion contrast using MCDI.



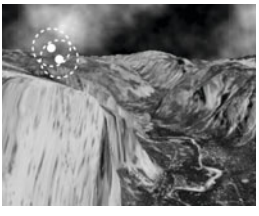
(a) Yosemite with Clouds (ground truth)



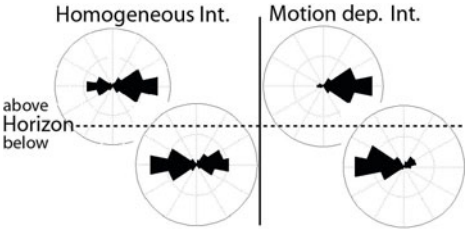
(b) motion contrast dependent integration



(c) context enhanced motion contrast



(d) Yosemite with Clouds (RF locations for (e))



(e) motion direction histogram (HI vs. MCDI)

Fig. 6. (a) The ground truth of the flow field of the 'Yosemite with Clouds' image sequence. (b) The estimated flow field using the MCDI mechanism. After the interpretation of the population code there is almost no improvement compared to the HI method. (c) MSTl responses show the estimated motion discontinuities used as integration boundaries. (d) Sample RF locations above and below the horizon with the corresponding motion direction histograms (e). Using the HI method ((e), left) the population code represents both, the cloud as well as the mountain movement independent of the RF position relative to the horizon. However, the MCDI method ((e), right) leads to a much more distinct representation of the underlying movements of the sky and the mountains.

where the horizon and the sky meet (Figure 6d). With HI, the population activity shows a multi-modal distribution since it represents both, the movement of the clouds to the right, as well as the movement of the mountains to the left (Figure 6e, left). This is because the RFs are covering regions above and below the horizon. For the population code interpretation (e.g. for flow field representations) the modus with the highest amplitude is chosen and the other modus, even though it is only slightly smaller, is ignored. However, using the MCDI method the movements of the clouds and the mountains are not mixed up, since the RFs are limited by the motion discontinuities (Figure 6c) and therefore avoid an overlap of the regions above and below the horizon. This results in an unimodal population code representation (Figure 6e, right) that offers a better source for further computations, e.g. the motion discontinuities in model area MST or the feedback from model area MT to V1. Furthermore the method for the interpretation of the population code can now be more general since the MCDI method prevents the emergence of ambiguous multi-modal representations.

4 Discussion and Conclusions

We propose a biologically inspired model of motion estimation that utilizes the low and intermediate stages of cortical motion processing, namely V1, MT and MSTl, to detect motion. Our model makes a new contribution regarding the process of motion integration at area MT. We propose that incorporation of motion discontinuities into the integration mechanism improves the result of motion estimation. Those activities yield a more reliable and more robust signal for segregation of moving objects. The resulting integration follows the Gestalt-law of common fate. In particular, this signal is exclusively gathered from the motion (dorsal) pathway and is not dependent on other features like luminance, form or texture. Since these features might not be visible or complicated to detect robust boundary estimation can be impossible. These channels might be integrated in future versions of the model in order to benefit from multi-feature contrast detection and fusion of boundary information.

At the moment, our scheme incorporates the response of motion contrast cells in MSTl that depict all changes in speed and direction. These noisy and fragmented estimates are improved with a methodology adapted from cortical areas V1 and V2, where lateral connections group together like-oriented contrasts and thus lead to contextually improved and less noisy signals. A diffusion process is then applied to adapt formerly circular integration areas to the outline of objects. Compared to classical homogeneous integration mechanisms our proposed motion contrast dependent integration scheme shows less estimation errors at object boundaries. The lower cortical region V1 benefits from this improved integration process due to recurrent feedback connection that stabilize and disambiguate the estimations. This property is clearly shown when multiple estimates are presented in velocity space. Here, our proposed extension leads to less ambivalent activations within cell populations.

Future work will focus on improving contextual enhancement and the estimation of motion discontinuities. Furthermore, classic features like texture statistics or signals derived from the form (ventral) pathway can be included in the estimation of object boundaries.

Acknowledgments and Author Contributions

This work was supported by a grant from the European Commission within the 7th Framework Program: Smart Eyes: Attending and Recognizing Instances of Salient Events (SEARISE, Project no. 215866). This work was also supported with a grant from the German Federal Ministry of Education and Research, project 01GW0763, Brain Plasticity and Perceptual Learning. Authors S. R. and S. T. have equally contributed to the contents of this publication.

References

1. Brox, T., Papenberg, N., Weickert, J.: High Accuracy Optical Flow Estimation Based on a Theory for Warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)

2. Physik, F.A.: Poggendorff's Annel (1855)
3. Horn, B., Schunck, B.: Determining optical Flow. Artificial Intelligence 17(1-3), 185–203 (1981)
4. Raudies, F., Neumann, H.: A model of neural mechanisms in monocular transparent motion perception. Journal of Physiology 104(1-2), 71–83 (2010)
5. Deqing, S., Roth, S., Black, M.: Secrets of Optical Flow Estimation and Their Principles. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2432–2439 (2010)
6. Tlapale, E., Masson, G., Kornprobst, P.: Modelling the dynamics of motion integration with a new luminance-gated diffusion mechanism. Vision Research 50(17), 1676–1692 (2010)
7. Weickert, J.: Anisotropic Diffusion in Image Processing. Teubner Verlag, Stuttgart (1998)
8. Eifuku, S., Wurtz, R.: Response to motion in extrastriate area MSTl: Disparity sensitivity. Journal of Neurophysiology 82(5), 2462 (1999)
9. Hegde, J., van Essen, D.: Selectivity for complex shapes in primate visual area V2. The Journal of Neuroscience 20, RC61 (2000)
10. Barron, J., Fleet, D., Beauchemin, S.: Performance of Optical Flow Techniques. International Journal of Computer Vision 12(1), 43–77 (1994)
11. Neumann, H., Sepp, W.: Recurrent V1-V2 interaction in early visual boundary. Processing. Biol. Cyber. 81, 425–444 (1999)
12. Thielscher, A., Neumann, H.: Neural mechanisms of human texture processing: texture boundary detection and visual search. Spatial vision 18(2), 227–257 (2005)
13. Weitzel, L., Kopecz, K., Spengler, C., Eckhorn, R., Reitboeck, H.: Contour Segmentation with Recurrent Neural Networks of Pulse-Coding Neurons. In: Int. Conference on Computer Analysis of Images and Patterns (2000)
14. Hansen, T., Neumann, H.: A recurrent model of contour integration in primary visual cortex. Journal of Vision 8, 1–25 (2008)