A MULTI-DIFFERENTIAL NEUROMORPHIC APPROACH TO MOTION DETECTION

Peter W. McOwan*, Christopher Benton, Jason Dale & Alan Johnston
*Department of Mathematical and Computing Sciences, Goldsmiths College, London SE14 6NW, UK
& Department of Psychology, University College London, London WC2 6BT,UK

This paper presents a multi-differential neuromorphic approach to motion detection. The model is based evidence for a differential operators interpretation of the properties of the cortical motion pathway. We discuss how this strategy, which provides a robust measure of speed for a range of types of image motion using a single computational mechanism, forms a useful framework in which to develop future neuromorphic motion systems. We also discuss both our approaches to developing computational motion models, and constraints in the design strategy for transferring motion models to other domains of early visual processing.

1. Nature as a Design Template for Neuromorphic Visual Systems

Nature, through the process of evolution, has developed efficient and robust strategies for the processing of environmental visual information. These algorithms have been optimised over aeons to maximise survival in the real world. One of the most fundamental visual sensory abilities that living biological systems require is the capacity to detect the motion of objects in the surrounding environment. These movements can be indicative of predator or prey, and allow the organism to chart its own movement through the world. It is not therefore surprising that there has been substantial effort devoted to developing motion-detecting chips based on biological models. Many of the current neuromorphic approaches to developing motion sensitive chips make use of models for insect visual systems. Specifically the correlation based approach of the Reichardt model has been used as a template in motion chip design [1,2] and incorporated into autonomous robots.[3]

The fly has obvious advantages as a model i.e. small size, tractable neural system and proven environmental functionality. We should however not limit our endeavors to emulate only fly vision, there are other interesting biological visual systems to copy. The fly correlation model does have a number of fundamental drawbacks. The basic correlation method is unable to function independent of the contrast of the moving object, a high output signal may be due to a high object contrast rather than a good spatio-temporal correlation. A single correlator is velocity tuned; it gives a peak response for a particular velocity that is fixed by the structural spatio-temporal parameters used. It is not therefore able to recover the actual velocity of an object moving at arbitrary speed. While these problems may not preclude the use of this model in certain specific applications, it lacks a general environmental robustness. By examining the cortical visual processing of higher animals, and man himself, we may be able to uncover mathematical principles that are more robust and could greatly enhance the sensor abilities of autonomous robots.

To discover these principles we can undertake psychophysical and neurophysiological studies of primates and formulate explicit mathematical models of the cortical motion pathway. This computational model provides a framework that permits the biologist to attribute mathematical functionality to neuro-anatomical structures and provides the psychologist with an understanding of the processes underpinning a subject's psychophysical performance. The robotics researcher gains an algorithm for the perceptual process, which can then be implemented in suitable hardware. Production of a mathematical model will also allow the generation of predictions that can then subsequently be tested experimentally. Feedback may be exploited to modify the model to better approximate the experimental neurophysiological and psychophysical data. With luck, this incremental modelling approach will converge on a realistic and biologically plausible solution, which a neuromorphic engineer may then subsequently exploit.

2. Design Principles in Space-time Models for Motion Perception

The space-time image is a device commonly used to represent the motion of a visual stimulus, and is an important aid to visualising the current competing models for motion perception. In this representation velocity becomes an orientation of the iso-brightness contours in the domain spanned by space and time. Psychophysical studies show that, for human observers, perceived velocity and the ability to discriminate small changes in velocity are largely unaffected by randomising the spatial and temporal frequencies of observed sinusoidal gratings so long as the physical speed remained constant.[4] This finding has been taken to imply that at some level within the motion processing scheme there is an explicit code for

velocity. The model we will present in this paper shows how this encoding may be achieved.

In line with a vast body of experimental data, all models of motion analysis in the mammalian visual system take as their starting point the existence of low-level spatio-temporal filters. However models can, and do, differ on the proposed strategy used by the visual system to both form these spatio-temporal filters, and the rules by which these filters are combined to compute motion. There are currently two main low-level approaches to the modelling of motion processing, Fourier Energy and Gradient methods, summarised in Figure 1.
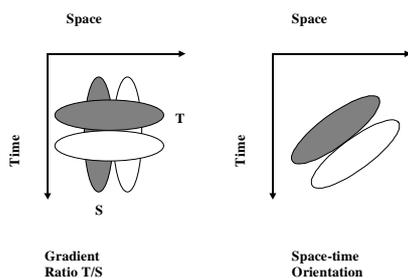

Figure 1. Gradient and Fourier energy models. In the gradient model the temporal derivative, shown as T in the above figure, is divided by the spatial derivative, S, to recover the speed. In the Fourier energy approach a bank of space-time oriented filters measure the orientation of the space-time structure.

In the Fourier Energy approach[5,6] the low level filters, often modelled as Gabor functions (a Gaussian multiplying an space-time oriented sinusoid), are combined to produce space-time oriented receptive fields, as shown in the right hand diagram in Figure 1.

` The responses from many different filters, each tuned to a specific spatio-temporal frequency, are then pooled and used to recover speed. The model in effect takes a measure of the orientation of the local Fourier power spectrum with each spatio-temporal filter. This model, though currently popular, suffers from the same drawbacks of sensitivity to contrast and filter velocity tuning as the Reichardt correlator from which it evolved. In addition where the stimulus power spectrum does not consist of components oriented about the frequency space origin, for example non-rigid motions or texture defined (second-order) motion, the basic Fourier energy model does not operate reliably, see (Ref.7).

In the Gradient approach the ratio of a temporal derivative filter T and the spatial derivative filter S is taken to recover speed, see left hand diagram, Figure 1. The filters are modelled as fuzzy derivatives (a Gaussian multiplied by a Hermite polynomial), see (Ref.8). This ratio technique is insensitive to the contrast of the stimulus, and gives a direct measure of stimulus speed.

We may distinguish between the biological validity of these two potential frameworks by examining how the measured cell profiles match the mathematical classification

of the low-level filters in the two models. When undertaking such a venture it is also important to examine both the biological plausibility of the mathematical operations used and the steps in data fitting through which any erroneous classification may occur. In the next section we present evidence for a derivatives based interpretation of the cortical motion pathway.

3. Modelling the Temporal and Spatial filters: Biological evidence for fuzzy derivatives

Three separate temporal channels that, in parallel, shape the time course of the visual signal have been identified psychophysically in humans using a masking paradigm.[9] The process of temporal differentiation can model accurately the relationships between the three experimentally measured temporal frequency sensitivity functions. Johnston and Clifford[10] show that the frequency domain descriptions of a causal, biologically plausible, Gaussian function of log time, and its first and second derivatives give an excellent fit to the psychophysical data. The log Gaussian has only two parameters to specify its shape.

In the spatial domain the form of the cell receptive fields in early visual cortex may be accurately modelled with blurred derivative of Gaussian functions.[11] As the order of differentiation increases, these derivative of Gaussian functions become tuned to higher spatial frequencies giving rise to the range of independent spatial channels found experimentally. Young also presents evidence that the distribution of zero-crossings of spatial receptive field shapes in monkey and cat are described better by Gaussian derivatives than by Gabor functions.[11] This physiological and psychophysical data would seem to provide evidence that the visual system implements a filtering process using fuzzy derivatives, evidence that supports the gradient based scheme. The fuzzy derivative model has only one parameter to fit, the variance of the underlying Gaussian. Gabor functions require two fitting parameters, the spread of the Gaussian and the spatial frequency of the underlying sinusoid and therefore may have a greater opportunity to fit the experimental data. All orders of the derivative receptive fields can be formed by the hierarchical superposition of spatially offset circularity symmetric Gaussian receptive fields, similar to those recorded in the LGN (Ref.11)

4. Accounting for Hierarchical Organisation and Non-linear Properties for Cortical Cells

To remain true to the philosophy of the neuromorphic field we must be able, thorough our computational algorithm, to account for recognised structures in the biological system we wish to copy. In addition to the low-level filters there are a number of well-established cell types in the visual cortex. These cells are often referred to as simple, complex

and hyper-complex or end stopped, based on their responses to moving gratings.[12] To simply have this catalogue of cell types is not, however to have a model of visual perception, in the same way as a zoo is not a model for natural selection. It is necessary to have a theoretical framework which can specify how these cell properties come about and understand how the cells combine to give rise to the perceived visual experience. We have demonstrated that the multiple measures model we present in this paper allows a unifying theory by which the properties of the known hierarchy of cortical cells may be described.[12,13] Additionally if we consider the non-linear properties of complex cells spatio-temporal receptive fields, which researchers measure using two-bar interaction techniques,[14] we find that our model gives the required space-time oriented interaction field measured physiologically.[15]

5. The Multiple Measures Model; Conditioning the basic gradient model

As illustrated in Figure 1 the simplest model that allows the recovery of speed from the space-time image structure is to form the ratio of the temporal to spatial derivative, as shown Eqn.

$$v = \frac{dx}{dt} = -\frac{\dfrac{\partial I(x,t)}{\partial t}}{\dfrac{\partial I(x,t)}{\partial x}} = -\frac{D_t I}{D_x I} \qquad (1)$$

(1). This is the basic gradient ratio model and recovers the speed independent of the stimulus contrast. Tanner & Mead have implemented a version of this mechanism in a motion chip, (see Ref. 16, Chapter 14). There is however a significant problem associated with this formulation; where the spatial derivative of the space-time images becomes zero the denominator vanishes. Division by zero is mathematically undefined, thus the simple gradient ratio model cannot recover the speed at all points in the stimulus. This computation is mathematically ill-conditioned.

$$v = \frac{\partial I}{\partial t} \Big/ \frac{\partial I}{\partial x}$$

$$v = \frac{\partial}{\partial x}\frac{\partial I}{\partial t} \Big/ \frac{\partial^2 I}{\partial x^2} \qquad (2)$$

$$\vdots$$

$$v = \frac{\partial^{n-1}}{\partial x^{n-1}}\frac{\partial I}{\partial t} \Big/ \frac{\partial^n I}{\partial x^n}$$

How then do we solve this mathematical problem? One pragmatic solution could be to apply a threshold to the calculation, evaluating the ratio only for a large enough value of the denominator or to spatially pool values of the spatial derivative in a local neighbourhood to form the denominator.[17] However it is known that the visual system measures at least three orders of spatial derivative,[11] and takes three orders of temporal differentiation.[10] In principle

therefore it would be possible to form low level filters to calculate the speed using different orders of differential operators, as given in Eqn. (2).

Through correctly balancing the ratio so that the numerator always carries one order of temporal differentiation more than the denominator, and the denominator carries one order of spatial differentiation more than the numerator, the cortex can have access to numerous measures of speed. These multiple measures of speed are still individually mathematically ill conditioned, with the denominator able to take on zero values at turning points on the space-time image. We may now, however, form two vectors X and T as shown in Eqn. (3), which contain the results of applying the fuzzy derivative operators to the image.

$$X = (\frac{\partial I}{\partial x}, \frac{\partial^2 I}{\partial x^2} \cdots \frac{\partial^n I}{\partial x^n})$$

In forming this higher dimensional

$$T = (\frac{\partial I}{\partial t}, \frac{\partial}{\partial x}\frac{\partial I}{\partial t} \cdots \frac{\partial^{n-1}}{\partial x^{n-1}}\frac{\partial I}{\partial t}) \qquad (3)$$

space we preserve the property that at each position within the vectors the term in X contains one more spatial derivative than in T, and in the T vector each term has one more order of temporal differentiation than the corresponding term in X.

To recover the best approximation to the speed from each of the measures we apply a least squares formulation to the vectors X and T, recovering a single value v'. The scalar product on the denominator of Eqn. (4), $X.X$, is now a sum of squares and therefore is always non-zero where there is spatial structure to measure. At maxima, minima and points of inflection in the space-time image the three orders of differentiation ensure that at least one term on the denominator is non -zero.[12] Substituting the local Taylor series expansion for the point on the space-time image[12] may further enhance this model. The model can also be extended to recover a robust measure of motion in two-dimensional images by combining well-conditioned measures of velocity and inverse velocity over a range of orientations, see (Ref.13) for

$$v' = \frac{X \cdot T}{X \cdot X}$$

$$v' = \frac{\sum_n \dfrac{\partial^n}{\partial x^n} I \dfrac{\partial^{n-1}}{\partial x^{n-1}}\dfrac{\partial}{\partial t} I}{\sum_n \dfrac{\partial^n}{\partial x^n} I \dfrac{\partial^n}{\partial x^n} I} \qquad (4)$$

details and a discussion of this model's resistance to image noise. Our approach is based on the general design strategy that to resolve mathematical ill-conditioning the human visual system takes multiple measures of the stimulus rather than to introduce arbitrary image dependant parameters such as thresholds. Such a multiple measures strategy both increases robustness, and when transferred to hardware, removes the requirement to tune or re-tune model

parameters during operation, so allowing truly autonomous operation.

5. Comparison with other Motion Analysis Methods and Application to Natural Image Sequences

Barron, Fleet and Beauchemin have undertaken a systematic analysis of the performance of a number of popular motion models.[18] We have also assessed the performance of the two dimensional version of our multiple-measures model[13] using Barron's performance metric and test motion sequence.

Figure 2 Recovered optical flow field for traffic sequence. The needle diagram illustrated shows the direction and speed of the motion at each point in the field. The length of the needle represents speed. A representative image from the traffic sequence used is shown inset at top right.

To give a comparable performance measure we follow Barron and calculate the angular error in space-time orientation between unit vectors in the direction of the correct image velocity vector v_c and the model estimated velocity vector v_e.

The angular error term $\psi_E = \arccos(\vec{v}_c \cdot \vec{v}_e)$ was calculated at each image position. Table 1 summarises the results for a synthetic image sequence used by Barron, a translating sinusoidal checkerboard (sine2). We also show results for the Horn and Schunck and Lucas and Kanade methods from (Ref.18).

It can be seen that, for this particular test motion sequence, our model performs better than the other two gradient methods, giving rise to lower average angular errors. We have also applied the multi-differential model to natural image sequences. The recovered velocity field for a traffic scene is shown in Figure 2. The model clearly picks up the movement of the two dark cars moving

rapidly past in the foreground, and the white car turning into the side street. This sequence suffers from real detector noise, yet the model can recover the motion and produce sharp motion boundaries without recourse to additional constraints, smoothing or iteration.

Table 1. Comparisons of motion algorithms, Average angular error and standard deviations (S.D.) shown for each method.

Method	Av. Ang. Error	S. D.
Lucas and Kanade	2.47	0.16
Horn and Schunck	0.97	2.62
Multi -Differential	0.40	0.43

6. A Design Principle for Neurmorphic Vision: Ockham's Razor and Second-order Motion

When designing neuromorphic sensor systems, for use in say autonomous robots, it is sensible to pursue an approach of minimum complexity. Invoking Ockham's razor `*It is vain to do with more what can be done with fewer'* (Ref.19) designers will tend to incorporate only the minimum set of features for the required functionality. There is no reason to assume that the process of biological evolution does not adhere to this philosophy, being driven by the same optimisation criteria as the engineer. When confronted with visual phenomena that cannot be accounted for by an established model, there is a tendency for researchers to postulate an additional visual channel to account specifically for this new data. We prefer to adhere to Ockham's razor and seek to explain the greatest range of psychophysical and neurophysiological results with a single model.

Currently the general view held is that the visual system operates using Fourier energy mechanisms (Figure 1) to recover the orientation of the Fourier power spectrum. This has been challenged by psychophysical investigations of second-order or texture based motion.[7]

A second-order motion stimulus, for an example stimulus see Figure 3, is constructed so that the Fourier components of the motion are not oriented through the frequency space origin. Examples of second-order motions are the movement of a contrast modulation over a field of random noise, or the movement of a contrast envelope over a moving sinusoid. Subjects observe the motion of second-order signals, though such motions are invisible to Fourier energy models, and hence to all neuromorphic systems based on this approach. It has been proposed that an additional, parallel non-linear motion pathway exists specifically for the recovery of second-order motion.[7] A rectifying front-end filter is added to the basic Fourier energy channel to demodulate the signal back to the frequency space origin and so allow the movement to be detected.

The full version of the model presented in Section 5 is able to extract the speed from a number of second-order motions. [12] We were able to predict correctly the perceived speed in the low contrast areas of a second-order contrast modulated sine grating, [20] see Figure 3, providing psychophysical validation for our model.
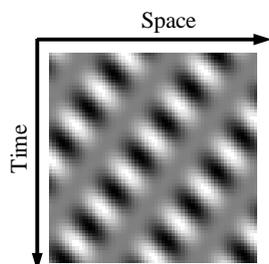
Space

Figure 3. The space-time image for a second-order motion: a contrast modulated sine grating. The contrast of the carrier grating is modulated by the second sine grating. In the stimulus shown the sinusoidal contrast envelope, the gray areas, moves to the left over time with the underlying sinusoidal carrier moving to the right. Both gratings are moving at the same speed but in opposite directions.

We are also able to explain a number of apparent motion illusions, [10] the perception of which has, in the past, been attributed to separate motion mechanisms. These abilities arise not through a specific design stage, but as rather as an emergent property of the non-linear processes in the model, the primary aim of which is to extract a robust measure of visual motion.

7. Towards a General Mathematical Framework for Visual Processing: The Plenoptic Function

If we have developed a successful approach for motion analysis, can we apply the same design strategy to other visual tasks? The input from the environment to the visual system can be represented as a multi-dimensional function of space, time, wavelength and viewpoint. [21] This plenoptic function contains all the information necessary to construct the basic measures of the visual process by computing measurements of orientation in this multi-dimensional space. What then is the mechanism by which the visual system recovers this orientation for a given sub domain of this space? It is common for researchers to model the different mechanisms by direct analogy. So, for example, Fourier energy models for motion perception construct space-time oriented receptive fields in direct analogy to the spatial orientation columns first reported in cortex by Hubel and Wiesel. [22] This analogical approach to model building is underpinned by the assumption that the mathematical formalism of the plenoptic function allows the computational principles for measuring orientation in the space-space domain to be simply transported to space-time to measure velocity, the orientation of structure in space-time. Such a notion has frequently underpinned theoretical

thinking for developing models in other perceptual domains for example stereo disparity. [23]

To test the validity of this transfer hypothesis, a psychophysical investigation comparing the pattern of results for a contrast modulated grating in space-time and space-space was undertaken. In the spatial task observers were asked to match the perceived spatial orientation of the low contrast regions of the modulated grating to that of a simple sine grating. In the motion analogue observers were asked to match the speed of the low contrast regions of the modulation to that of a moving sine grating.

The pattern of results can be summarised as follows, Figure 4. In the spatial domain orientation was perceived veridically when carrier and envelope orientation were parallel or orthogonal. The largest error in perceived orientation, which could be as much as 10^0, was produced when the carrier was vertical. [24] In the space-time domain the same stimulus gave a veridical speed match only when carrier and envelope were parallel, as the carrier orientation (speed of the carrier grating) moved towards the orthogonal condition observers perceived a slowing down of the low contrast regions. When carrier and envelope were exactly orthogonal observers perceived the low contrast regions to be stationary. [10]
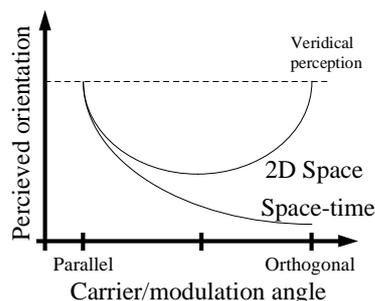

Figure 4. An illustration of the two distinct forms of dependency of perceived orientation of the low contrast regions of the modulation envelope as a function of carrier orientation for 2D space and space-time, see (Ref. 24).

The evidence from this psychophysical comparison reveals that the biological system uses separate, distinct strategies for encoding orientation in two-dimensional space and space-time, so invalidating the simple transfer of models hypothesis. In formulating valid models for the different visual modalities it is necessary to customise the mechanism using domain specific properties.

8. Conclusions

In this paper we have presented the principles we feel are important when developing a biologically plausible mathematical framework for the detection of motion. We have argued that neuromorphic vision systems using the multiple differential measure technique have the potential to

mimic the performance of the human visual pathway and deal with a range of types of motion. Additionally, we have indicated the potential advantages in constructing robust front-end processors for autonomous systems using this design strategy and demonstrated the model's performance with real motion sequences. Finally, we discussed the potential to transfer this multi-measure computational strategy from the motion domain to other allied visual domains, and indicated the usefulness of psychophysical experiments to ascertain the validity and limits on such transfers.

Acknowledgements

References

1. Delbrook T., Silicon retina with correlation based velocity tuned pixels, *IEEE Trans. Neural networks* **4**, 529-541. (1993)

2. Arias-Estrada M., Tremblay M. & Poussart D., A focal plane architecture for motion computation. *Real-Time Imaging* **2,** 351-360. (1996)

3. Francheshini N., Pichon J. M. & Blanes C., From insect vision to robot vision*, Phil. Trans R. Soc. Lond.* B **337,** 283-294. (1992)

4. McKee S. P., Silverman G. H. & Nakayama, K. Precise velocity discrimination despite random variations in temporal frequency and contrast. *Vision Research*, **26**, 609-619. (1986)

5. Heeger, D. J., Model for the extraction of image flow. *J. Opt. Soc. Am. A.* **4**, 1455-1471. (1987)

6. Watson, A. B. & Ahumada, A. J., Model of human visual-motion sensing. *J. Opt. Soc. Am. A.* **2**, 322-341. (1985)

7. Chubb, C. & Sperling, G., Drift balanced random dot stimuli; a general basis for studying non Fourier motion. *J. Opt. Soc. Am. A.* **5**, 1986-2007. (1988)

8. Koenderink J. J. & Van Doorn A. J., Representations of local geometry in the visual system, *Biol. Cybern.* **55**, 367-375. (1987)

9. Hess R. F. & Snowden R. J., Temporal properties of human visual filters: Number, shapes and spatial covariance, *Vision Research.* **32**, 47-60. (1992)

10. Johnston A. & Clifford C. W. G. A unified account of three apparent motion illusions, *Vision Research*, **35**, 1109-1123. (1995)

11. Young R.. The Gaussian derivative theory of spatial vision: analysis of cortical cell receptive field weighting profiles. *General Motors Research Report* GMR 4920. (1985)

12. Johnston, A., McOwan, P.W. & Buxton, H.. A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc. R. Soc. Lond. B.* **250**, 297-306. (1992)

13. Johnston A., McOwan P. W. & Benton C., Robust velocity computation from a biologically motivated model of motion perception *Proc. Royal Soc. Lond. B* **266**, 509-518 (1999)

14. Emerson R. C., Bergen J. R. & Adelson E. H., Directionally selective complex cells and the computation of motion energy in cat visual cortex, *Vision Research.* **32**, 203-218. (1992)

15. Johnston A., McOwan P. W. & Benton C. Non-linear interaction fields in directionally selective complex cell receptive fields are predicted by a gradient motion model, *Invest. Opthal. & Vis. Sci* **36,** 277, (1995)

16. Mead, C., Analogue VLSI and neural systems, Addison-Wesley Publishing Co, Reading MA (1989)

17. Lucas B. D. & Kanade T., An iterative algorithm registration technique with a application to stereo vision *Proc 7th Int Joint Conf on AI* Vancouver B.C., 674-679. (1981)

18. Barron J.L., Fleet D.J. & Beauchemin S. S. Performance of Optical Flow Techniques*, International Journal of Computer Vision* **12** , 43-77 (1994)

19. Russell B., *History of western philosophy and its connection with political and social circumstances from the earliest times to the present day.* Allen & Unwin, London. (1946)

20. Johnston A. & Clifford C. W. G. Perceived motion of contrast modulated gratings: predictions of the Multi-channel Gradient model and the role of full wave rectification, *Vision Research*, **35**, 1771-1783. (1995)

21. Adelson E. H. & Bergen J. R, The plenoptic function and the elements of early vision, In *Computational Models of early vision*, MIT Press, 2-20. (1991)

22. Hubel D. H. & Wiesel T. N., Functional architecture of macaque monkey visual cortex (Ferrier Lecture) *Proc. R. Soc Lond B,* **198,** 1-59 (1977)

23. Ohzawa I., DeAngelis G. C. & Freeman R. D. Stereoscopic depth discrimination in the visual cortex, neurons ideally suited as disparity detectors *Science* **249,** 1037-1041 (1990)

24. McOwan P. W. & Johnston A., A second-order pattern reveals separate strategies for encoding orientation in two-dimensional space and space-time, *Vision Research*, **36**, 425-430. (1996)