EXTRACTION OF HIGHER ORDER DERIVATIVES OF THE OPTICAL VELOCITY VECTOR FIELD: LIMITATIONS IMPOSED BY BIOLOGICAL HARDWARE

Ken Nakayama. Ph.D.

Smith-Kettlewell Institute of Visual Sciences
Medical Research Institute
2200 Webster Street, San Francisco, CA 94115

In the past two decades there has been an increasing awareness of the biological importance of optical flow fields. This work has proceeded along three rather distinct paths: theoretical, psychophysical and physiological.

Theoretical work on this subject began with Helmholtz who noted the difference in optical velocity for different points of the visual field corresponding to objects at different distances (2). Thus began the use of the term motion parallax as a potential cue to depth (3). Gibson elaborated this idea in several ways. First, he emphasized the existence of higher order invariant aspects of this vector field, noting how they carried useful information, including the layout of surfaces in depth. Later mathematical papers have shown that various linear differential operators such as deformation, vorticity, dilation, and acceleration contain rich and often explicit information, including the layout of surfaces, the rigidity of objects, and properties of observer motion (4, 5, 6). In addition, Nakayama and Loomis (7) proposed a plausible neural operator which would carry information regarding surface boundaries. This latter proposal is summarized in Figures 1 and 2.

In terms of psychophysical studies, two sets of experiments stand out. First was the demonstration of the kinetic depth effect by Wallach and O'Connell (8). They demonstrated that a simple two dimensional shadow of a 3 dimensional figure can appear in vivid depth when the figure is rotated. Under most circumstances, however, the depth percept remains ambiguous such that near and far points are often seen in reversed depth. More decisive was the recent discovery by Rogers and Graham
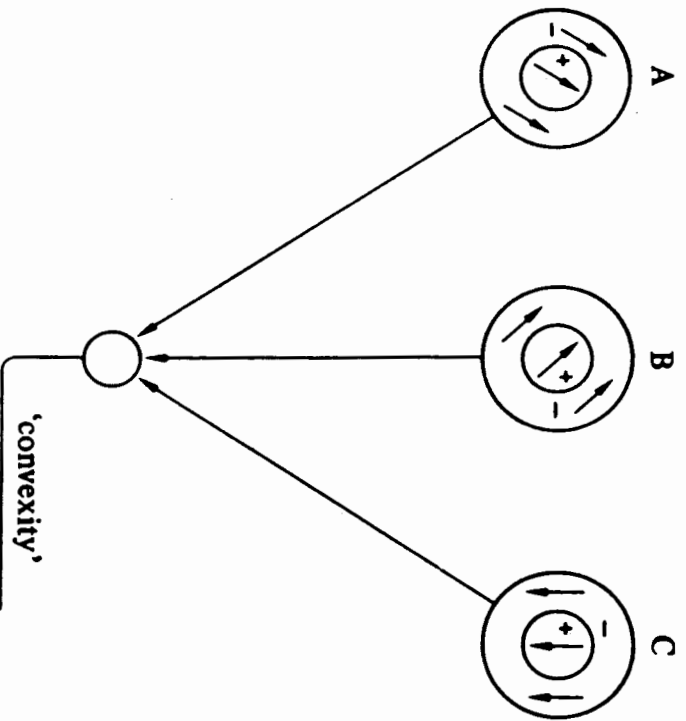
Figure 1. Physiological wiring diagram of the "convexity" operator proposed to extract edges of three dimensional objects for an observer translating through a rigid environment. The convexity unit consists of subunits comprised of Kuffler-type receptive fields summing center/surround velocity signals antagonistically. Each subunit is centered on the same retinal locus and analyzes a particular direction of motion. These subunits are then summed in an excitatory fashion to obtain a "convexity" value for all portions of the visual field. Such a function is insensitive to uniform motion in a given retinal region and will only respond when there are velocity discontinuities between neighboring regions in the visual field (7).
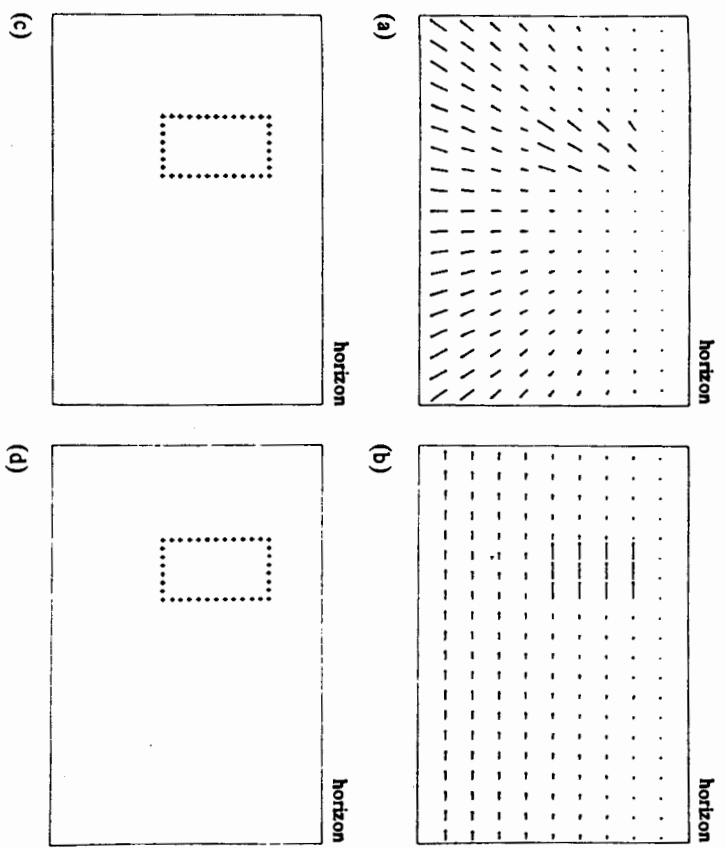

Figure 2. Illustrative example of how the convexity operator can delineate the boundaries of real three dimensional objects. In A and B, we simulate the same set of external environmental features, consisting of a ground plane and a flat fronto-parallel rectangular surface suspended over this ground plane. In A, the observer is translating into the plane of the paper, in B he is translating to the right. Note that the optical flow field looks very different in each case. The convexity function of each flow field, as represented by C and D, however, is essentially the same, delineating the edge of the rectangular screen under very different conditions. Thus, the convexity operator can encode visual boundaries of 3-dimensional objects independent of changes in the local direction of the motion vectors (7).

(9) of a totally unambiguous depth percept from differential motion of random dots. Pure depth without motion can be seen despite the existence of differential retinal motion created by their display. Thus Rogers and Graham's monocular pattern is essentially indistinguishable from a random dot stereogram (10). See Rogers (this volume).

Physiological recordings from single neurons show that responses to moving configurations are rather similar to theoretical mechanisms proposed to code significant features of the optical flow field. Frost and Nakayama (11) found that all units in the deeper layers of the pigeon optic tectum were organized in a manner very similar to the "convexity" operator as shown in Figure 1. These cells responded to opposing motion of a center spot with respect to a background independent of the direction of movement. As expected, these cells did not respond to whole field motion and this was further confirmed by deoxyglucose mapping techniques (see Frost, this volume). Other examples of the detection of relative motion by single units can be seen in the cat striate cortex (12; see also Hammond this volume) and in MT cortex of monkey (13).

The remarkable convergence of these separate approaches (theoretical, psychophysical, and physiological) suggests that the vector field concept as originally proposed by Gibson and elaborated more mathematically by others, might be regarded as a promising approach. The flow field contains higher order variables which are highly informative about the layout of objects in the three dimensional world and the observer's own movements. Humans actually see depth in flow fields. Finally, neuro-physiological mechanisms exist which could code higher order variables of the optic array.

Given this success it might seem inappropriate to be critical of this general line of thinking, especially as it provides an exciting new way to understand a wide variety of phenomena. At this point, it would also seem important, however, to evaluate this approach in terms of newly emerging information regarding the biological hardware characteristics of early motion processing. Therefore, the major purpose of this paper is to examine the degree to which the basic biological hardware of the visual system might impose important constraints on how optical flow fields might be encoded, thus providing some limits on future mathematical theorizing.

Before proceeding we will recapitulate the general line of thinking introduced by Gibson and followed by others, namely, the vector field approach to the optical flow field. In this view,

it is assumed that the vector field of optical velocities contains very useful information and that it is encoded directly and that higher order invariants of this field are derived. Some of the information available by considering higher order invariants is summarized in Table I.

Table I

Theoretical information available from higher order variables of the optical flow field.

Information	Higher order invariant or operator	Reference
1. Slant of surface	Deformation	Koenderink & van Doorn (4)
2. Gaussian curvature		" (4)
3. Rigidity of surface	Second temporal derivative	Longuet-Higgins & Prazdny (6)
4. Time to collision	Dilation	Lee & Reddish (14)
5. Visual boundaries of 3-D objects	Convexity	Nakayama & Loomis (7)

In theory, it might seem sufficient to compute these mathematical aspects of the velocity field and very useful information would emerge, information that would aid greatly in the organisms' survival.

We need to remind ourselves at this point, however, that these ideas are mathematical and physical concepts, not biological ones. Obviously the existence of information in the optic array is no guarantee that it will be used by ourselves or other animals. As an analogy, consider the job faced by the astronomer or other "observer" who has historically relied on data from the static and changing optic array. From this he has measured the distances of the stars and their atomic composition. All of this very significant information about object properties is clearly available in the optic array, yet our own visual system cannot encode it without the use of specialized optical instruments (spectrographs, telescopes, etc). We are biological creatures, evolved to obtain particular sorts of information and limited by the capacities of our optical and neurosensory systems. In this paper I suggest the existence of at least three limitations imposed by the properties of our motion sensing system.

1. SPATIAL BANDWIDTH LIMITATIONS

The visual system encodes luminance variations to only a finite level of spatial detail. Thus the well known contrast sensitivity function in humans falls off above 6 cycles/degree in the fovea and at progressively lower values in the periphery (16). With differential shearing motion, however, there appears to be a reduction in sensitivity above 0.6 cycles/degree; ten times lower than for luminance sensitivity (15). (See Figure 3). The lack of spatial resolution means that the motion threshold is
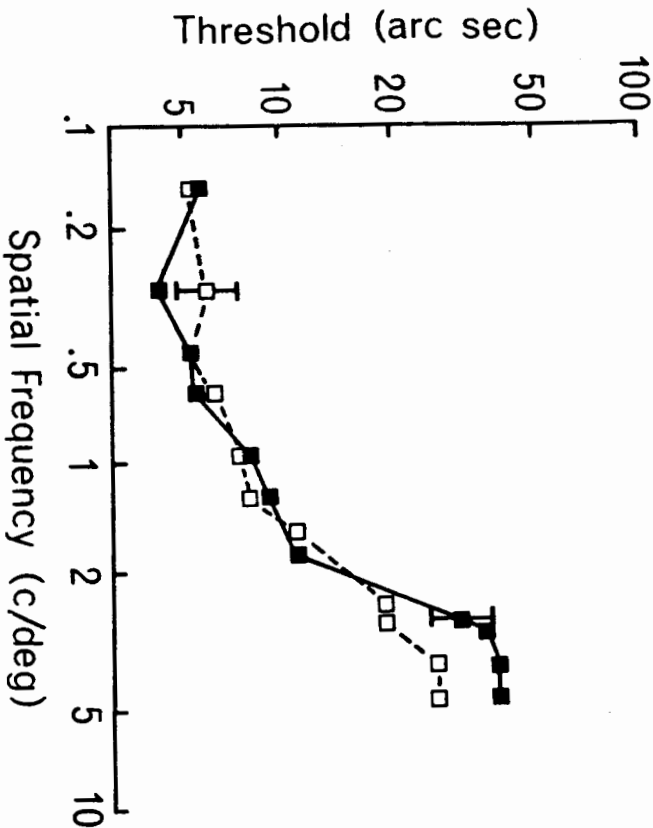

Figure 3. Spatial bandwith limitations for motion sensitivity. Shearing motion thresholds as a function of the spatial frequency of the shearing motion. Stimulus is a screen full of random dots where the horizontal shearing motion is a sinusoidal function of vertical position. Note that thresholds rise above around 0.6 cycles/degree. Temporal frequency is 2 Hz (15).

not based on the velocities of individual points but rather on their mean value integrated over a large distance, perhaps by as much as 15' in the foveal area. Evidence also suggests that the integration of motion is anisotropic, having summation properties greater in a direction orthogonal to the local motion vector rather than parallel to it (17). The spatial bandwidth for compressive motion is about twice the bandwidth for shearing motion. The existence of a rather large elongated receptive field for the encoding of motion even in the fovea suggests higher-order consequences. It implies that any hypothetical topographic map of velocities must be significantly smeared and more so in the direction orthogonal to the local motion vector than parallel to it.

In terms of the higher order encoding of the velocity field, it suggests that the mathematical operation of taking a spatial derivative of the motion field is somewhat misleading. More likely, the system first integrates the velocity signal spatially, then takes spatial derivatives. This was implied by Nakayama and Loomis (7) in their definition of the convexity operator and explicitly formulated mathematically in the luminance domain by Marr and Hildreth (18). In practice, this means that any information derived from the optical flow field is spatially very coarse.

This coarseness is clearly seen in the results of Rogers and Graham (9) where the appreciation of depth from motion is very reduced above the very low spatial frequency of 0.5 cycles/degree, even though the appreciation of fine luminance detail is reduced only for patterns ten times as fine.

Our spatial resolution is limited more in depth than in luminance. For example, in a dense forest scene we have sufficient luminance resolution to resolve individual leaves on trees but because of the lack of spatial resolution in the motion system, we cannot sense their individual depth values. We see depth, but more coarsely, more at the level of clumps or branches containing many leaves. A strikingly similar lack of depth resolution was originally described for stereopsis (19).

This lack of spatial resolution is not necessarily a biological handicap. On the contrary, the low resolution of the motion system would be advantageous, providing useful information without requiring a reduplication of yet another high resolution representation, and its concomitant need for a large number of neurons. A similar situation exists in the encoding of chromaticity, especially pronounced in the blue cone pathway. Blue cones are extremely sparse yet they provide a vivid percept of color, one that is an extremely useful adjunct in the process of image segmentation.

2. TEMPORAL BANDWIDTH LIMITATIONS

The visual system is limited temporally as well as spatially. The modulation sensitivity for flicker falls off above 10-20 Hz depending on the brightness and configuration of the stimulus (20, 21). The temporal bandwidth of the motion system is much lower. This can be measured psychophysically by presenting a shearing motion in random dots and varying temporal frequency. If one plots the threshold peak velocity of the sinusoidal temporal stimulus, it rises rapidly above 1 Hz (see Figure 4). This indicates that motion information is temporally integrated, a property embodied in many previous models of motion processing (22,23,24).
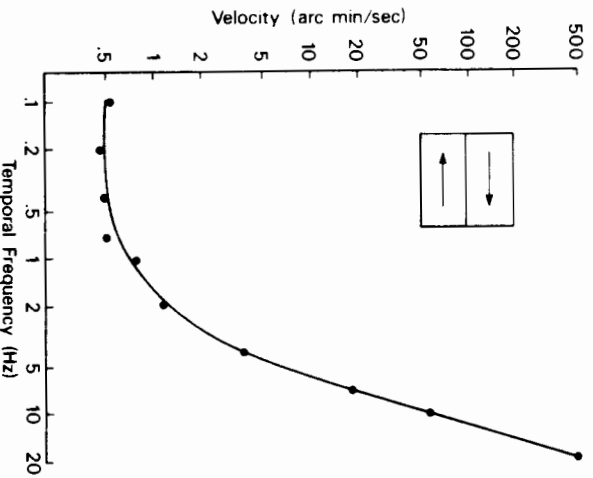

Figure 4. Temporal bandwidth limitations for motion sensitivity. Shearing velocity thresholds as a function of temporal frequency of sinusoidal oscillation of random dots. Note that thresholds rise above 1 Hz. (Replotted from similar data described in reference 15, Figure 3.)

What are the implications of this lack of temporal resolution in motion sensitivity? There are at least two and they are related. First we should note that this property of temporal low-pass filtering puts a rather severe limitation on the visual system's ability to sense acceleration. Thus mathematical descriptions of the acceleration field (25) as well as computational theories to derive information from this field are likely to have limited applicability. As an example, the checking of object rigidity by calculating the acceleration may prove too slow to have functional value (6). Empirical work where acceleration sensitivity is measured confirms this view. Acceleration is poorly detected (26).

In addition. the low temporal frequency sensitivity of the visual motion system also casts doubts on theories or formulations requiring 3 views in close succession. Ullman's (27) structure from motion theorem would seem to have reduced biological applicability insofar as his theory implicitly requires the pick-up of 2 different velocity snapshots in quick succession, a property severely limited by the low-pass characteristics of motion sensitivity.

3. LIMITATIONS IMPOSED BY NON-LINEARITIES (WEBER'S LAW FOR VELOCITY)

Up until now all mathematical theories suggesting mechanisms to extract information from velocity vector fields have employed linear differential operators. Thus the curl of the velocity field can be defined independent of other terms, such as divergence. or common image motion. As an illustrative example, consider how the system might compute the gradient of the velocity field (Gibson's original proposal), obtaining a 2x2 matrix summarizing how the x and y components of the velocity field vary in the x and y direction (5).

$$\text{Grad } V = \begin{bmatrix} \dfrac{\partial V_x}{\partial X} & \dfrac{\partial V_x}{\partial Y} \\[2ex] \dfrac{\partial V_y}{\partial X} & \dfrac{\partial V_y}{\partial Y} \end{bmatrix} \qquad (1)$$

where V_x, V_y are the x and y components of V. Suppose that we were to add a common motion vector C in a given neighborhood, perhaps resulting as a consequence of an eye movement. Then

mathematically

$$\text{Grad.}(V + C) = \text{Grad } V \qquad (2)$$

because the constant term C has a derivative of zero.

As a consequence of Weber's law, however, one cannot assume the nervous system can obtain an unvarying Grad V for various degrees of added common image motion (C). Most neuronal signals are proportional to the adapting level of stimulation and as a corollary the nervous system is generally unsuited for the accurate differencing of large signals independent of signal level. Thus, it might not be expected to compute the terms of Grad (V + C), especially for large values of C. Empirical evidence supports this completely. By measuring the sensitivity to shearing motion in random dots, Nakayama (28) was able to show a progressive rise in differential shearing thresholds even for rather small amounts of common motion, (C). The relation followed Weber's law, having a Weber fraction of about .05. A similar adherence to Weber's Law was also seen for a wide range of suprathreshold velocity differences (29).

The existence of Weber's law for velocity suggests that the nervous system cannot make accurate measurements of Grad V or its components, nor can it compute any other linear differential operation. As a consequence, all mathematical theories at least those proposed so far, have limited applicability. We must recognize that the nervous system can extract biologically relevant information but that it cannot do it in the form specifically described by the simple differencing operation implied by taking a spatial derivative. Two alternatives deserve consideration: first is the possibility of other algorithms, taking the ratio of velocities of neighboring regions. This is consistent with the data and has some computational advantages (see ref. 7). Second one might question the need for highly metrical data at all. Perhaps ordinal relations are sufficient to obtain the needed information.

CONCLUSIONS

Very useful information is theoretically available from the vector field of optical velocities and humans respond to an array of differentially moving points by seeing depth unambiguously. Furthermore, it is possible that physiological mechanisms can code important higher order aspects of the velocity field and could provide the basis for image segmentation and depth discrimination.

The analysis of the moving optic array, however, is constrained by the hardware characteristics of the neurosensory encoding system. In particular, the spatial and temporal resolution of the movement system severely limits the amount of potential information that can be obtained from the optical velocity vector field, suggesting that the information provided is spatially very coarse and temporally very slow.

Finally, the existence of a Weber's law non-linearity for velocity limits the applicability of current mathematical theories which extract higher order spatial derivatives of the optical flow field. Other approaches based on taking ratios of neighboring velocities or even non-metrical schemes might be considered.

REFERENCES

1. Gibson, J.J. The perception of the Visual World. Houghton-Mifflin Co. (1950).

2. Helmholtz, H. Treatise on physiological optics. Translated by J.P. Southall, 1925, New York, Dover (1910).

3. Graham. C.H., K.E. Baker. M. Hecht, V.V. Lloyd. Factors influencing thresholds for monocular movement parallax. J. Exp. Psychol. 38 (1948) 205-223.

4. Koenderink, J.J. and A.J. van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. Optica Acta 22 (1975) 773-791.

5. Koenderink, J.J. and A.J. van Doorn. Local structure of movement parallax of the plane. J.Opt.Soc. Am. 66 (1976), 717-723.

6. Longuet-Higgins, H.C. and K. Prazdny. The interpretation of moving retinal images. Proc. R. Soc. Lond., B 208 (1980), 385-387.

7. Nakayama. K. and J.M. Loomis. Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. Perception 3 (1974), 63-80.

8. Wallach, H. and D.N. O'Connell. The kinetic depth effect. J. Exp. Psychol. 45 (1953), 205-217.

9. Rogers, B. and M. Graham. Motion parallax as an independent cue for depth perception. Perception 8 (1979), 125-134.

10. Julesz, B. Binocular depth perception of computer generated patterns. Bell System Technical Journal 39 (1960), 1125-1162.

11. Frost, B.J. and K. Nakayama. Single visual neurons code opposing motion independent of direction. Science 220 (1983), 744-745.

12. Hammond, P. and A.T. Smith. On the sensitivity of complex cells in feline striate cortex to relative motion. Exp. Brain Res. 47 (1982), 457-460.

13. Miezin, F., E. McGuiness and J. Allman. Antagonistic direction specific mechanisms in area MT in the owl monkey. Neuroscience Abstracts 8 (1982) 681.

14. Lee, D.N. and P.E. Reddish. Plummeting gannets: a paradigm of ecological optics. Nature 293 (1980), 293-294.

15. Nakayama. K. and C.W. Tyler. Psychophysical isolation of motion sensitivity by removal of familiar position cues. Vision Res. 21 (1981), 427-433.

16. Virsu, V. and J. Rovamo. Visual resolution, contrast sensitivity, and the cortical magnification factor. Exp. Brain Res. 37 (1979), 475-494.

17. Nakayama. K., G. Silverman, J. Mulligan and D.I.A. MacLeod. Anisotropic spatial integration in the perception of motion. Perception 11, (1982), A6.

18. Marr, D. and E. Hildreth. Theory of edge detection. Proc. R. Soc. Lond. B 207 (1980), 187-217.

19. Tyler, C.W. Spatial organization of binocular disparity sensitivity. Vision Res. 15 (1975), 583-590.

20. de Lange, H. Attenuation characteristics and phase characteristics of the human fovea-cortex systems in relation to flicker fusion phenomena. Doctoral dissertation. Technische Hogeschool, Delft (1957).

21. Kelly, D.H. Visual responses to time-dependent stimuli. I. Amplitude sensitivity measurements. J. Opt. Soc. Am. 51 (1961), 422-429.

22. Reichardt, W. Autocorrelation: a principle for the evaluation of sensory information by the central nervous system. In, Sensory Communication (Ed. W.A. Rosenblith), (1961), New York, Wiley, 303-318.

23. Foster. D.C. A model of the human visual system in its responses to certain classes of moving stimuli. Kybernetik 8 (1971), 69-84.

24. Marr. D. and S. Ullman. Directional selectivity and its use in early visual processing. Proc. R. Soc. Lond., B 211 (1981), 151-180.

25. Gordon, D.A. Static and dynamic visual fields in human space perception. J. Opt. Soc. Am. 55 (1965), 1296-1303.

26. Gottsdanker, R.M. The ability of human operators to detect acceleration of target motion. Psychological Bulletin 53 (1956), 477-487.

27. Ullman, S. The interpretation of structure from motion. Proc. R .Soc. Lond., B 203 (1979), 405-426.

28. Nakayama K. Differential motion hyperacuity under conditions of common image motion. Vision Res. 21 (1981), 1475-1482.

29. Van Doorn, A.J. and J.J. Koenderink. Visibility of movement gradients. Biological Cybernetics 44 (1982), 167-175.