

OPTIC FLOW

JAN J KOENDERINK

Department of Medical and Physiological Physics, Physics Laboratory, State University of Utrecht,
Princetonplein 5, 3584 CC Utrecht, The Netherlands

Abstract—This paper offers a quick review of the subject of “optic flow” in its conceptual and computational aspects. The theory is evaluated in terms of possible applications in the neurophysiology and experimental psychology of spatial sensorymotor behaviour and perception. The problem of which kind of detector is suited to extract various aspects of optic flow is given special attention. It is shown that the possibilities are actually much more various than is reflected in the current (even the frankly speculative) literature. It is argued that a system that is sensitive to the relative time changes of the orientation differences of image details is especially suited for an analysis of the optic flow with regard to the information concerning the three dimensional shape of objects such as is contained in the flow. Thus the orientation sensitive elements that are known to be abundantly present in the primary visual cortex of many vertebrates are hereby implicated as a quite likely substrate for the extraction of the solid shape of environmental objects. In our opinion this possibility should be investigated with the same ardour as the usual interpretation, which holds this system responsible for the initial extraction of the contours of flat (i.e. defined in the image) shapes. A new, partial solution to the “structure from motion problem” is offered, that not only covers the usual case of shape extraction in the presence of rigid motions of the object, but also the much wider class of (non-rigid) bending deformations (such as occur in the non-rigid deformations of inextensible shells). These solutions violate all conditions required by the well known “structure from motion theorem” the solutions are possible for point configurations in which no fourtuple of points moves as a rigid structure and for input data from merely two views. A numerical example illustrates how this algorithm can be used to predict side views of an object from very limited input data.

Optic flow Motion parallax Solid shape Rigid motion Non-rigid motion Depth illusions

1. THE NOTION OF “OPTIC FLOW”

1.1 Historical

A clear understanding of the fact that the deformation of the retinal image due to ego-motion or the transposition of objects in the environment is not just a nuisance but actually a rich source of information concerning the world had already been reached by such scientists as Hermann von Helmholtz (1910) and Ernst Mach (1886) in the 19th century. However, they did not probe very deep into the theory of the problem. This is surprising in view of the fact that nothing else but Euclidean geometry is needed to do the basic analysis: all the tools were there. Modern developments start with James Gibson (1950) who, not being trained in mathematics, does not seem to have had access to the tools needed for a basic analysis but who displayed the intuitive sense of a real genius for what is important about a problem. Although he made quite a few slips it seems fair to say that he pointed out about everything that seems worthwhile to study in the subject and that modern developments generally

follow in his footsteps (That is apart from his curious notion of “direct perception”. Ullman, 1980). Modern developments have been predominantly inspired by practical needs. Flow field analysis has been used in the study of the control of movement of various vehicles in road and air traffic (Gibson *et al.*, 1958; Gordon, 1965; Kruk and Regan, 1983) and nowadays developments come from AI and robotics (Koenderink and van Doorn, 1975, 1976a, b, 1978, 1981, 1984; Ullman, 1979, Regan *et al.*, 1979; Prazdny, 1980; Longuet-Higgins and Prazdny, 1981; Longuet-Higgins, 1981).

It may be useful to sketch Gibson's early approach to the problem. As a perceptual psychologist concerned with ergonomic problems he was highly interested in depth perception in the sense of how observers estimate their position and orientation with respect to their environment, how they estimate parameters needed to plan a course through the environment, etc. As such he was thoroughly familiar with the classical “depth cues”, and in fact added one himself, the “texture gradient” cue. (Gibson, 1950. Although the cue had been used by artists

for centuries I'm not sure that it was considered important in science before Gibson made so much of it.) Given a well defined "grain-size" of detail in the world (e.g. pebbles of—at least statistically—similar size, pavement tiles, grasses, bushes, etc.), it can be inferred from simple optics that this is translated into a texture in the retinal image that is not even but instead is space variant: the retinal grain size diminishes with increasing distance to the eye. In many circumstances this can be a powerful cue (Stevens, 1981). It is a dangerous one, however: e.g. a painter can easily fool us with painted texture gradients (Of course this is equally true for the other static depth cues.) On a pavement on which the size of the tiles is subtly modulated with distance you can also fall prey to depth "illusions". Such possibilities have been exploited by the late Renaissance artists with effects that are still able to stun us. This is because the actual grain size enters in the retinal image: the texture gradient does not depend merely on slope and distance but also on the nature of the objects in the world. Gibson saw that this problem does not arise with the motion parallax gradients present in the spatiotemporal deformations of the retinal image that are entirely due to ego-motion: indeed the motion parallax depends only on slope and distance, not at all on the nature of the objects "out there". The objects (or rather their images) merely serve as "tags" or "tokens" that enable us to extract the optic flow which carries them along. Thus the cue of motion-parallax-gradient offers objective information about the movement and the lay-out of the surroundings irrespective of the precise nature of the environment. Moreover, the active observer (during locomotion) actually *controls* the flow and thus objectifies his information in a sense that is impossible in the static case: a painted texture gradient leads to an optic flow that is characteristic for the plane canvas, not for the imaginary space suggested by the painter. This fact is exactly what makes optic flow information of interest to the robotics community. (Of course we are not saying that the visual system will always use the available information to full advantage, in fact exceptions are known. But this is a case for psychophysics, not for the present theoretical exposition.)

1.2. Local vs global features of the flow

It is useful to distinguish from the start between *local* and *global* features of the optic

flow. Let me consider some important global features first.

One global feature that is extremely common concerns the flow induced by a rotation around an axis through the vantage point itself to a good approximation (neglecting the effect that the first nodal point does not coincide with the momentary center of rotation of the eye): eye-movements fall into this class. It will be useful to introduce the notion (due to Gibson) of the "optic array" here: the optic array is the two dimensional manifold of visual directions. Although there exists no particular mathematical reason to do so it is often convenient (for the sake of discussion, or as an aid to intuition, etc.) to think of the optic array as a sphere centered around the vantage point (You may consider such a sphere to be just a parametrization of the optic array, providing us with spherical coordinates. No actual projection is implied here. Thus the remark that the optic array is an approximation because the eye-lens is on the surface of "the" sphere completely misses the point.) Then rotations around an axis through the vantage point induce rigid rotations of the optic array. This is an optic flow that depends only on the movement and not at all on the spatial lay-out of the environment. Thus you can't gain *depth* information from the flow by making eye movements.

Another important global feature is induced through a *translation* of the vantage point. The pattern of streamlines induced by a pure translation (movement along a straight line) does not depend on the structure of the environment either, but only on the movement. This pattern consists of a family of longitude circles on the optic array: the flow is from one pole along the longitude circles to the other one. (The poles are commonly known as vanishing points.) A few precautions are necessary. First of all it is strictly wrong to speak of the pattern of *streamlines*. As in hydrodynamics we distinguish *streamlines* [which are the orbits traced out by "material" points (or tokens)], *streaklines* [which are the orbits of material points that once occupied a given place (e.g. the smoke plume of a chimney is the standard hydrodynamical example)], and *fieldlines* (which are the integral curves of the momentaneous velocity field, i.e. at all points the velocity vector is tangent to the fieldlines). In the general dynamic case all these curves are different and the global feature considered here applies only to the fieldlines of the optic flow. The "optic flows"

which are commonly depicted by comic book artists generally appear to be inspired on stream and streak lines, thus the observation does not apply to them. Another fact that needs to be pointed out is that the value of the velocity along the fieldlines is not part of this global feature: it may vary irregularly from point to point and it does depend on the spatial lay-out of the environment whereas the pattern of fieldlines does not.

A third global feature that is of recurring importance is somewhat more complicated. It depends on the fact that the environment is largely composed of three-dimensional bodies of which only the generally smooth (on the levels of resolution considered here) *surfaces* can imprint themselves on the optic flow. (At least for the case of optics such a statement is an apt description of the natural world because most objects tend to be opaque.) Now smooth surfaces induce smooth flows, thus the optic flow will consist of piece-wise smooth regions. At the boundaries of such regions we must generally expect discontinuities. Within the regions the flow is smoothly varying, but generally not constant: it varies from place to place.

Local features are of two types: the first is the average flow velocity at that locality. The second is the structure of the local variation of velocity in the immediate neighbourhood of the locality, it is also known as the *motion parallax* field. Motion parallax is important because it does not suppose an absolute direction as reference. Local variations lead to deformation of image detail. e.g. a small square drawn on the optic array and subjected to the flow transforms into a parallelogram, etc. The mathematical analysis suited to describe the motion parallax is deformation analysis. The basic results are simple enough (Koenderink and van Doorn, 1975).

Any small deformation (and the deformation can be made as small as you please by regarding very small patches of the optic array and/or regarding only short time spans during which the patch is subjected to the flow) leads to a linear (affine) transformation of the patch that can be decomposed into four basic components [these components are themselves transformations—albeit of a simple kind—and can be added again (the order is immaterial) to regain the original complicated transformation].

a *translation* (which does not deform at all!), an *isotropic expansion* or contraction (“homothety” which strictly speaking also does not deform the visual field locally, but merely imposes a scaling factor), a *rigid rotation* (which is of course a local isometry, and thus also does not impose a true deformation) and a *pure shear* (a contraction in one and an expansion in the orthogonal direction, such that area is conserved). The latter three will be referred to as the *dw*, *curl* and *def* components. These are “differential invariants” of the flow: they do not depend on the choice of the coordinate system, which at once explains their potential importance for organic systems. The *dw* is a number that specifies the relative time change of apparent area (solid angle) of a piece of the optic array, the *curl* is a number that specifies the rate of rotation, and the *def* can be specified with a number (the degree of shear always positive) and an orientation (the axis of contraction say).

The gradients of the *dw*, *curl* and *def* are the entities that are of prime interest as “cues”. Before I discuss them we have to specify how the differential invariants depend on the spatial structure of the environment and on the parameters describing the motion. In other words we have to introduce *dynamics* as opposed to the pure *kinematics* described in this section.

1.3. Dynamics of the optic flow

1.3.1. The movement parameters. It should be obvious that only the relative movement of observer and environment matters as far as optic flow is concerned. In the case that we are concerned with a generic local patch of the optic array—for which we may expect a smooth flow—we may as well attribute the relative movement to the observer*. This relative movement may be described in various ways. However, one method is especially convenient in the treatment of optic flow: to simplify our problem we may invoke Chasles’ theorem (Whittaker, 1904) which states that each movement can be decomposed into a translation and a rotation around an axis through a prescribed point in a unique manner. Let the (instantaneous) rotation axis be through the vantage point. Then the effect of the rotation is a trivial rigid rotation of the optic array (which carries no exterospecific information at all!) and the only nontrivial component left is the translation. Let the translational velocity be denoted \mathbf{V} , the rotational velocity \mathbf{R} . Let the surface patch we are inter-

*Such an essentially arbitrary choice serves to make the mathematical analysis much more concise and clear.

ested in be at a direction \mathbf{r} (unit vector in space, or point of the optic array) Then it makes sense to decompose the vectors \mathbf{V} and \mathbf{R} into *radial* components (V_r, R_r), which are the components in the direction of \mathbf{r} and *transverse* components (V_t, R_t), which are the components perpendicular to \mathbf{r} . V_r and R_r may be treated as scalar fields on the optic array, V_t and R_t as "tangent vectors" (Two-dimensional vectors defined on the manifold of visual directions because we know that the radial components of V_t and R_t vanishes identically (by definition) this is possible) Even better, we may introduce the distance to the patch (d) and define the "specific radial velocity" $A_r = V_r/d$ and the "specific transverse velocity" $A_t = V_t/d$ these entities have the dimension (time)⁻¹ just as R_r, R_t have In this way all spatial units are excluded from the very beginning This is both convenient and important because it is *a priori* clear that all flow field quantities will have to be expressed in terms of A_r, A_t, R_r and R_t . [Because only quantities of the dimension of (time) are available in the flow, it makes sense to express everything in terms of quantities of this dimension.]

These quantities have an obvious intuitive content: R_r is a rotation around the line of sight, R_t describes a "pan" or "tilt" of the eye. A_r has an interesting interpretation. it is the reciprocal of the time needed to reach the patch at distance d with the radial velocity V_r . Thus it may well be called "immediacy", or nearness in time It is often known as the inverse of the "time to collision" (Lee, 1976, 1980), although this strictly only applies for the case that the transverse component vanishes The quantity A_t can be interpreted as an "apparent rotation" It measures the local and instantaneous rate of turn in the optic array due to translation.

1.3.2. *The parameters specifying surface layout* A surface patch may be described (relative to the vantage point) with increasing accuracy in the following manner

- its distance (d) is specified (0th order description),
- its orientation is specified (1st order description the tangent plane is thereby specified) (Stevens, 1983a, b);
- its curvatures are specified (2nd order description, e.g. convexity or concavity is determined now),
- and so one can go on, adding term after term. [Technically speaking this is merely a Taylor expansion of the dis-

tance (or equivalently the nearness) around a given visual direction]

The differential invariants div, curl and def can be specified in terms of 1st order description For their gradients the 2nd order terms are necessary. Thus I concentrate here on the 1st order terms (0th order term being trivial)

The orientation of the patch has a vector character it has a magnitude (or *slant*) as well as a direction (or *tilt*) For instance, one may use the angle between the inward normal to the patch and the visual direction as a measure of the slant, whereas the direction of increasing distance specifies the tilt, which may be treated conveniently as a tangent vector in the optic array A similar result is obtained by considering the tangent vector $\mathbf{F} = \text{grad} \log(d_0/d)$ the gradient of the reciprocal distance (or nearness) as measured in the optic array. The constant d_0 has no influence on the result and shows the basic fact of scale independence of orientation especially well. The magnitude of \mathbf{F} measures the tangent of the slant (vanishes for pure "frontal" view, is infinite for pure "side" view of the patch), whereas its direction is in the direction of fastest increase of the nearness (thus specifying tilt) This representation is most convenient in practice

1.3.3. *The differential invariants expressed in the basic parameters of motion and surface position and orientation* After the basic definitions of the previous sections it is a simple matter to describe the local flow field in terms of the basic parameters. First of all the local average translation is

$$\mathbf{t} = -\mathbf{A}_r - \mathbf{R}_t.$$

In this formula $-\mathbf{A}_r$ measures the classical disparity induced by a movement, whereas $-\mathbf{R}_t$ describes the shift due to a refixation (shift of fixation point). It is clear that for any movement we may choose to annul \mathbf{t} with a suitable smooth eye-movement (merely choose $\mathbf{R}_t = -\mathbf{A}_r$), but this can only work locally (because \mathbf{A}_r depends on distance) It may be useful in practice because a high modulus of the translation may lead to blur (Whiteside and Samuel, 1970).

The 1st order differential invariants depend also on the transverse components. For instance the def equals the product of the moduli of \mathbf{F} and \mathbf{A}_t , whereas the axis of contraction is the bisectrix of \mathbf{F} and \mathbf{A}_t . Thus the deformation does not depend on the radial components at all! This will prove to have important practical consequences (*vide infra*) The div and curl both

show two terms, one due to the radial and one due to the transverse components of the movement I shall discuss these separately

A translation towards a surface patch obviously leads to an expansion of its image (a positive div), whereas the reverse movement leads to a contraction (a negative div) Thus we obtain a term $2A_r$ in the div.

Similar a rotation in clockwise direction around the line of sight leads to a reverse rotation of the image, leading to a term $-2R_r$ in the curl

The transverse components lead to less trivial results. The contributions to div and curl depend on the relative orientations of \mathbf{F} and \mathbf{A}_r , for instance the contribution to the div vanishes if \mathbf{F} and \mathbf{A}_r are perpendicular, whereas the contribution to the curl vanishes if they are of the same direction This is most conveniently expressed in terms of the two products

- $\mathbf{F} \cdot \mathbf{A}_r$ (“dot” product, or “inner” product), which equals the product of the moduli of \mathbf{F} and \mathbf{A}_r times the cosine of the angle between them,
- $\mathbf{F} \times \mathbf{A}_r$ (“cross” product, or “outer” product), which equals the product of the moduli of \mathbf{F} and \mathbf{A}_r times the sine of the angle between them reckoned in the clockwise sense from \mathbf{F} to \mathbf{A}_r .

The complete formulae are:

$$\text{div} = -\mathbf{F} \cdot \mathbf{A}_r + 2A_r,$$

$$\text{curl} = -\mathbf{F} \times \mathbf{A}_r - 2R_r,$$

$$\text{def} = FA_r,$$

axis of contraction bisects \mathbf{F} and \mathbf{A}_r .

A formal derivation may be found in Koenderink and van Doorn (1975), but the intuitive content of the equations is so clear as to make the derivation almost superfluous (Figs 1 and 2). The basic facts can be nicely demonstrated with a “gauge figure” that allows one to estimate the nature of the deformations easily. In one example I use a square checkerboard pattern for this purpose [Fig. 1(a)].

Panning or tilting the camera results in a translation (no deformation) of the image [Figure 1(d) shows the result of a “pan”.] A similar result is obtained when the camera is moved sideways or up and down because the gauge figure is the image of a patch that is perpendicular to the optical axis the dot-product $\mathbf{F} \cdot \mathbf{A}_r$ vanishes A rotation of the camera

around the optical axis results in a rigid rotation of the image [Fig 1(c)] This illustrates the term $2R_r$ in the curl equation. A movement of the camera along the optical axis results in a mere size change. [Figure 1(b) shows a movement towards the object.] This illustrates the term $2A_r$ in the div equation.

The previous results were so trivial because the surface patch was not slanted. I now introduce a slant, but in such a devious manner that we still have a useful gauge figure in the image [Fig. 2(a, b)]. The method is simple I use not a checkerboard pattern, but a *predeformed* one. This pattern looks like a trapezoid with unequal checks when seen in the frontal view [Fig. 2(a)] However, when this object is rotated over 45° around the vertical axis it looks like a square checkerboard pattern [Fig 2(b)] (at least when the distance is right) It is still possible to detect the trick in the image: note that one of the hands of the person holding the object is imaged much larger than the other one Of course this is the side that is turned towards the camera Thus the nearness increases from right to left in the image, the vector \mathbf{F} is directed horizontally towards the left

With this set-up it is easy to demonstrate the effect of the remaining terms. We take care to move the camera in such a way that all radial components of the movement vanish. Panning and tilting are used to keep the gauge figure within the field of view: they have no influence on the deformations anyway.

First note what happens when the camera is moved towards the left [Fig. 2(c), the transformations in the image of the assistant allow one to spot the type of camera movement easily.] In that case the vectors \mathbf{A}_r and \mathbf{F} both point towards the left, thus their bisectrix is horizontal. Hence the axis of contraction should be horizontal. Note that this proves to be the actual outcome. Moreover, the curl should vanish (the cross product is zero) which also is evident from the result, and the divergence should be negative indeed the area of the gauge figure in the image has decreased. In a similar way a movement to the right results in an \mathbf{A}_r vector to the right and an \mathbf{F} vector to the left, thus a bisectrix in the vertical direction. Then the contraction is along the vertical, the divergence positive (net area increase) whereas the curl again vanishes Some reflection makes these results appear almost trivial For a movement towards the right the object is seen in a more frontal view, for a movement towards the left in

a less frontal view. This at once explains the phenomena, the content of the basic formulae just sums up the geometry in an especially orderly and useful manner

Finally note the result of a camera movement in the upward direction [Fig. 2(d)]. Because A_x points upwards and F towards the left, the axis of contraction should be in the oblique direction (45° with the vertical) from the lower-right to the upper-left. The experiment yields exactly this result. The square gauge figure has become a parallelogram with the same vertical sides as the original square: thus area is conserved, hence the divergence vanishes. (This figures with the formula because the dot product vanishes.) There is a net rotation though because the cross product does not vanish, but is positive: thus there is a counterclockwise rotation associated with this motion. This is noticeable because the vertical sides of the gauge figure have remained vertical whereas a pure shear would have turned them clockwise: thus there must be an additional counterclockwise rotation.

The basic set of equations captures all these cases (and in fact any combination of them). It sums up the geometry in an especially convenient fashion

1.3.4. Which kind of mechanisms are needed to extract the flow parameters? I don't want to go into too much detail here, especially I want to avoid discussion concerning such moot issues as "the correspondence problem", or "the aperture problem". Let us then assume that local velocities can be measured in some way or other and that image features can be compared with respect to size, position, and orientation, both simultaneously and successively. Granted such possibilities (which themselves appear to pose formidable research problems!) how can the flow parameters be derived?

The answer must differ for the global and the local entities

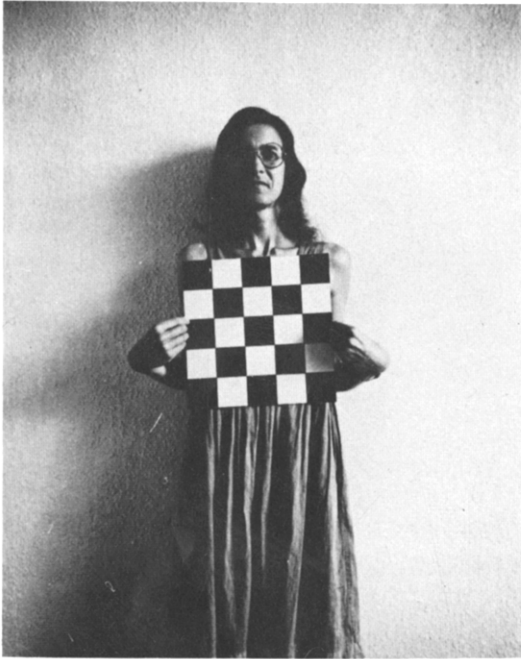
A measure of *global rotation* can be obtained by calculation of the total moment of the velocities in the optic array around three mutually perpendicular axes. This boils down to three weighted integrations. Implementation of such mechanisms seems simple and could easily be imagined to exist in physiologically acceptable "wetware", this would at once provide you with an estimate of R . Such mechanisms appear to have been demonstrated electrophysiologically (Simpson *et al.*, 1981).

A measure of *global translation* is more difficult to obtain because the velocity in the optic array induced by a movement depends on distance. It is possible to derive estimates of the vanishing points* from the directions, however (Perhaps making suitable use of the magnitudes to set weights.) The problem can be shown to be equivalent to that of finding the "best common point" of a set of straight lines in the plane. It is the same problem as that which was solved by the astronomers in the previous century in order to derive the proper motion of the sun among the stars of which the parallactic motions had been measured. Possible indications for the presence of such mechanisms in organic vision have been reported (Berthoz *et al.*, 1975, Cynader and Regan, 1978; Regan *et al.*, 1979, 1983). Once you have found the vanishing points, you have found the *direction* of the translation. The *magnitude* cannot be extracted without prior knowledge of distances. If the (in many cases quite reasonable) hypothesis that the world is on the average at rest is ventured, then you are in a position to extract the *relative acceleration* from the flow. (The linear acceleration divided by the linear velocity.) This possibility has to the best of my knowledge never been explored.

It is often possible to obtain an estimate of A_x , *locally*. This is the case because the influence of rotation does not depend on distance at all. Thus at an object boundary (when you are apt to find a depth transient) the transient in the optic flow must be due to the difference in A_x alone. There are other possibilities for a curved surface patch: it is possible to find A_x from the gradients of the div and curl for instance (Koenderink and van Doorn, 1975).

The extraction of differential invariants can take place in a multitude of ways. The possibility stressed most often (Koenderink and van Doorn, 1975, 1978, Longuet-Higgins and Prazdny, 1980) utilizes local velocity measurements. Indeed, the differential invariants can easily be expressed in terms of partial derivatives of the components of the velocity in some convenient coordinate system. Using a discrete approximation to the derivative at once yields a possible implementation for the div, def and curl (Koenderink and van Doorn, 1978). This is so obvious that this possibility should not tempt us to disregard other possibilities! For instance, there exist integral theorems which express the average value of the differential invariants in terms of integrals of the velocity around the

*That is: the sources and the sinks of the flowfield



(a)



(b)



(c)



(d)

Fig 1 A quick demonstration of the practical consequences of camera (or eye!) movements for the deformations of the image. The gauge figure in this case is a plane, square, checkered board. It is held perpendicular to the camera axis by the assistant. (a) A pure frontal view. The camera axis pierces the center of the object. Compare all other pictures to this fiducial image. (b) A camera translation along the camera axis towards the object. Result: a homogeneous expansion of the gauge figure, thus positive div, vanishing curl and def components. (c) A camera rotation around the camera axis. Result: a counter-rotation of the image, thus a curl, and vanishing div and def components. (d) A translation of the camera to the right. The result of this motion is the same as that of a camera rotation around a vertical axis to the right and in fact can be cancelled by such a rotation. All differential invariants vanish: the result is a pure translation in the (local) optic array.

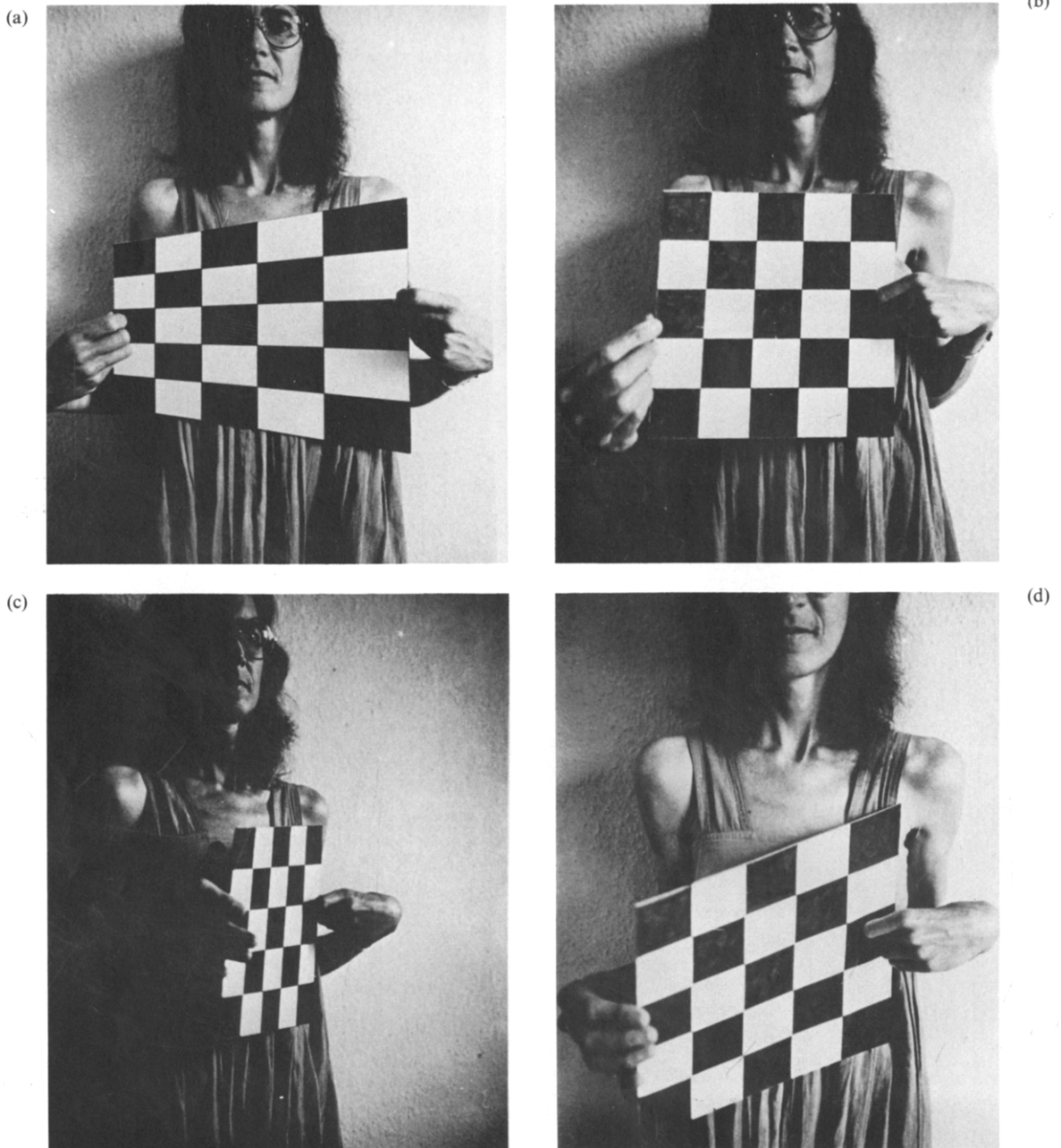


Fig 2 In this figure the results due to basic camera movements in the case of a surface element that is slanted with respect to the camera axis are demonstrated Camera rotations around the nodal point (panning and tilting) are freely used to keep the gauge figure in the image frame these have no consequences for the differential invariants anyway In this case I distinguish between the object in the hands of the assistant (which is a predeformed "square" of trapezoidal shape) and the gauge figure in the visual field (which happens to look square because of the curious shape of the object and the specific slant and tilt that is being applied to it) This makes the gauge figure an easy target for comparison (a) A frontal view of the object the assistant holds it plane perpendicular to the camera axis (b) The object held in its intended position it now looks "square" It is obvious from the photograph that the left hand of the assistant is more distant than the right one In fact the object was turned over 45° around a vertical axis The nearness gradient (on the photograph!) points to the left (c) The effect of a translation of the camera towards the left The bisectrix of the nearness gradient and the movement is horizontal, as is the axis of contraction Note that the area of the image of the gauge figure is decreased (compare c, d to b!), whereas there is no appreciable rotation (d) The camera was translated upwards (look at the hands to verify this) The bisectrix of movement and nearness gradient now runs from the upper left to the lower right of the photograph (45° slant) as does the axis of contraction Note that the area of the gauge figure has not changed, but that there is a definite rotation as a consequence to this movement You should figure out for yourself what the effect of opposite movements is [Answer movement to the right axis of contraction vertical, expansion, no curl, movement downwards axis of contraction from lower left to upper right (under 45°), no area change, definite curl]

boundary of regions. Such theorems suggest other possibilities, and in fact they often provide the basis for estimates of the differential invariants of various physical fields in engineering settings. Such methods are especially nice when singularities can occur, because they deal only with averages, not with point properties. Still other possibilities are suggested by general *conservation* laws. For instance, temporal texture density changes must be due to local divergence and this is expressed in an "equation of continuity". Thus monitoring texture density changes can yield estimates of the div.

There exists one possible method to obtain the shear component that seems especially suited for implementation in a physiological setting (*vide infra*). This method employs the changes in the orientations of textures elements. It is best illustrated with a few examples (Figs 3 and 4). This time I use a spoked wheel as a gauge figure. Note what happens with it when

we subject it to various transformations (Fig 3). A pure divergence just magnifies the gauge figure, thus it does not change the orientations of the spokes. Hence if you monitor orientations you automatically discard the div component. A pure curl changes the orientations of all spokes by the same amount. Thus the angles between two spokes are left invariant by the curl. Hence if we monitor orientation differences we automatically discard the curl component too. A pure deformation also changes orientation differences, although it does not change the average orientation. One easily shows that relative orientation difference changes (the change in the angle subtended by two successive spokes divided by this angle) uniquely specify the shear, both as far as magnitude and as far as orientation are concerned. Figures 3 and 4 show this in detail. One of the insets of Fig. 3 illustrates a general deformation (finite values for div, curl and def). In Fig 4 (upper graph) the angular

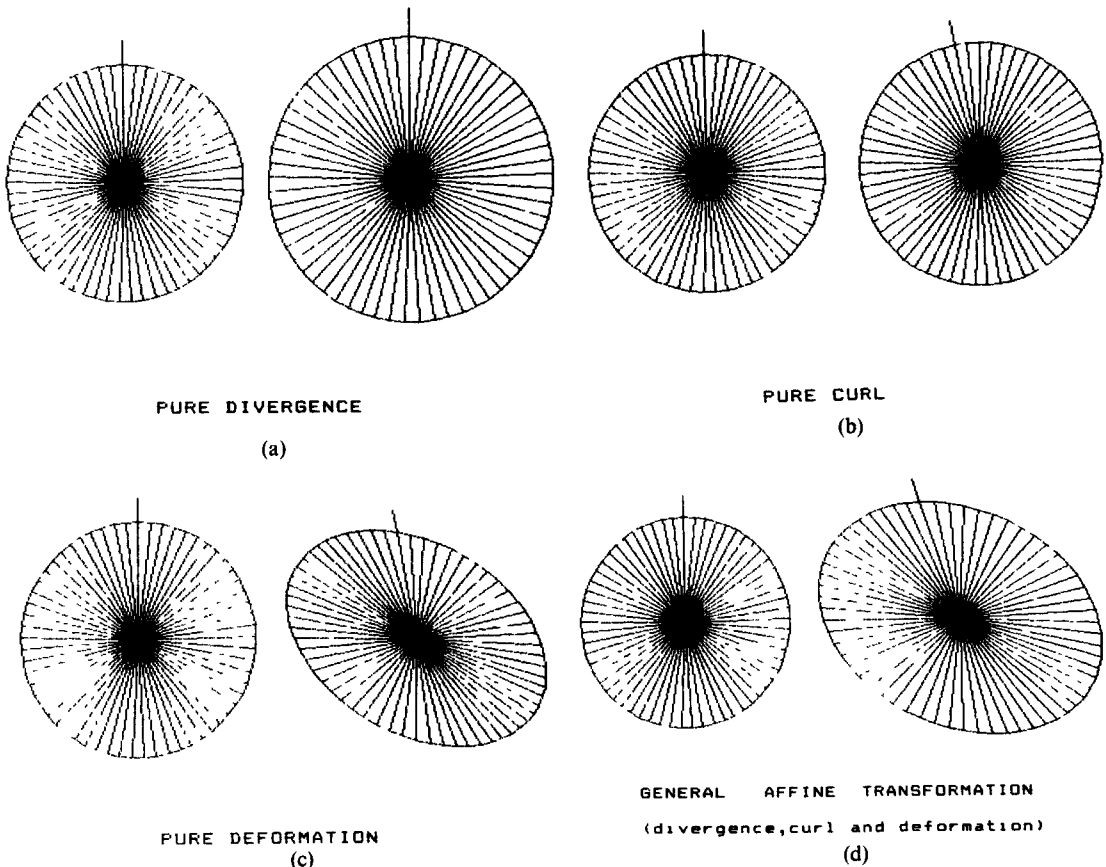


Fig 3 The changes in a simple gauge figure (spoked wheel) due to simple optic flow fields, namely a pure divergence, a pure curl and a pure shear. The final inset shows the result of a combination. Note especially in all figures —the area change (showing the effect of the div), —the net rotation (showing the effect of the curl, it can be estimated from the orientation of the single prolonged spoke, which gives a reference. The net rotation is the average orientation change over all spokes), —the change in the angle between successive spokes as a function of orientation (this feature is due to the def)

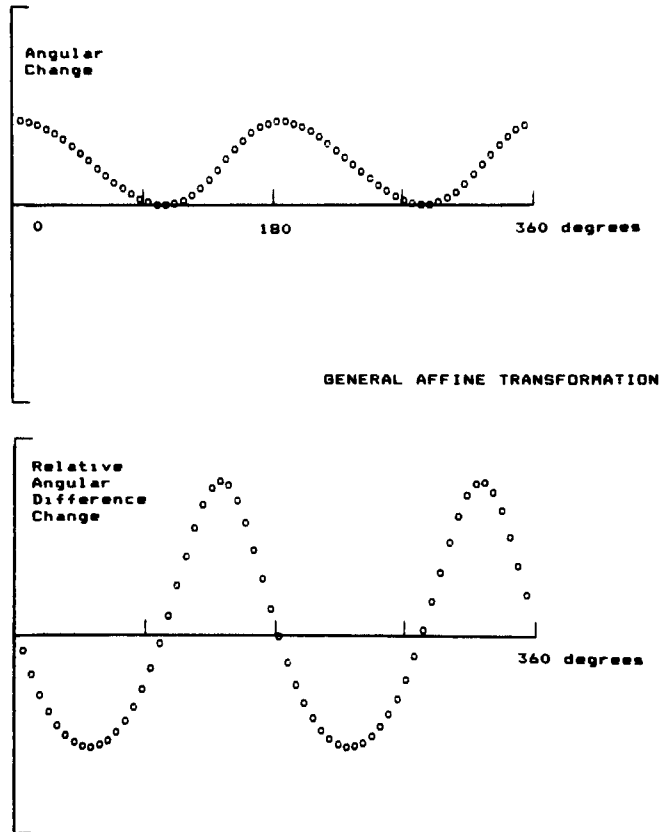


Fig 4 This graph shows results for the general affine deformation depicted in the previous figure *Upper graph* the angular change of a spoke as a function of its orientation Note the d c-shift due to the curl, and the a c-ripple due to the def *Lower graph* the difference of the angles between successive spokes divided by their sum is plotted as a function of orientation This result is completely independent of div and curl and depends merely on the def component The amplitude is proportional to the magnitude of the def, the phase defines the orientation of the axis of contraction

change of a spoke is plotted against orientation for this deformation. The “d c-component” is due to the curl, the modulation to the def. All div information is suppressed. In Fig 4 (lower graph) I plot relative angular difference change as a function of orientation. Note that the d.c.-component has disappeared. this graph depends only on the shear component. The amplitude of the curve specifies the strength of the shear, the zeros (or extrema) specify its direction

The point to keep in mind is that there is no need whatsoever to base the analysis of optic flow on velocity measurements: many alternative implementations (several with better noise immunity) exist This fact is generally ignored in the literature.

1.4. Information contained in optic flows

The optic flow contains information of various kinds. In order to put some structure on the

discussion I shall distinguish the following types:

- proprioceptive (in a purely visual sense!) information about ego-motion; both rotational and translational movements;
- information that is useful to sustain egocentric orientation and localization,
- information concerning the segmentation of the visual field into coherent entities and of the visual world into coherent (i.e. suffering “common fate”) objects. Thus the optic flow sustains segmentation, as well as aggregation (both “splitting” and “merging” in the usual jargon of computer science);
- exteroceptive information concerning the spatial structure of the surroundings, including relative motion of objects and nonrigid deformations

In the literature these items have received very

unequal attention, moreover discussions have been paralyzed or directed very one-sidedly because of preoccupations with the so called "rigidity hypothesis", or with exact solutions of the "structure from motion problem" (which is usually formulated in terms of minimum requirements in such a way as to exclude useful alternative routes from the outset). In the sequel I will discuss these concepts, then discuss the four points, two of them at the hand of an example.

1.4.1. The nature of various assumptions and approximations and the value of partial solutions. In the theoretical treatment of the interpretation of optic flow one generally invokes several assumptions, some of them explicitly, others silently, but all of them equally essential. Some of them are:

- continuous velocity fields are due to smooth surfaces in motion;
- changes in the flow are due to smooth transformations in 3-space

One usually invokes either:

- the "rigidity hypothesis". the smooth transformations are actually Euclidean isometries, i.e. they conserve mutual distances globally (Koenderink and van Doorn, 1975, 1976a, b, 1978, 1981; Ullman, 1979; Longuet-Higgins and Prazdny, 1980; Longuet-Higgins, 1981),

or:

- the "local rigidity hypothesis". the smooth transformations are actually Euclidean bendings, i.e. they conserve mutual distances measured along the surface, but not necessarily globally throughout space (this paper);

or.

- the "piecewise rigidity hypothesis" pieces of the object move as rigid objects, the transition parts that glue them together may suffer arbitrary deformations in 3-space, e.g. both bendings and stretchings these are then ignored (Hoffman and Flinchbaugh, 1982; Todd, 1982),
- when a velocity field does stop abruptly at a boundary, then the object in 3-space is being occluded, if the field ends in a graceful manner it is itself the occluder

The first assumption appears to be absolutely necessary and has to be relied on blindly. Some form of rigidity constraint is also necessary, although the strongest version can certainly be relaxed as will be shown later in an example. The piecewise rigidity assumption is the one usually employed by draftsmen of e.g. the human torso.

Although the simple rigidity assumption appears to be a strong one, it is actually quite reasonable in many circumstances (Mach, 1886). Whenever the movements of the observer can be considered fast on the time scale of typical environmental changes, then the whole environment can be regarded as one huge rigid body. This is very common and must have helped in shaping the visual system throughout organic evolution. The fact that an optical stimulation with violently deforming images can induce a strong impression of rigidity in space never fails to amaze us, the result is extremely powerful (Metzger, 1934, Braunstein, 1962). In case the observer entertains invalid prior hypotheses the apparent violent and lawful deformations of objectively rigid objects is equally striking (Mach, 1886; Rosenberg, 1924).

The class of bending deformations appears to be the largest one for which useful solutions of the "structure from motion problem" may be obtained. Transformations including stretchings are obviously too general for instance the optic flow on the face of a TV-tube includes stretchings. We often perceive apparent 3-D shapes on TV, yet objectively considered the transformations should yield only the flat CRT tube as a solution: but these are certainly not evoked by this flow! The typical formulation of the "structure from motion theorem" (Ullman, 1979) excludes the case of bendings. It states that it takes at least three views of at least four points (in general position) in order to compute the 3-D configuration. Yet I will show in the sequel that it is possible to find configurations of (say) seven points, no four of which move together in a rigid fashion, for which useful partial solutions can be obtained from only two views.

This brings me to the issue of the relative value of partial versus "complete" solutions. I intend to treat this matter in a pragmatic fashion in any specific case one should use the method that leads most easily to the actually required result. In many cases a partial solution is all we need and a quick partial solution is to be preferred over a cumbersome complete one.

which also yields data that are going to be ignored anyway. This is especially true if a complete solution can be obtained simply from the results of repeated partial ones, which often proves to be the case. If one is concerned with models for vision, one should also consider the possibility that this system does not compute complete solutions at all, but instead a number of quick and robust partial ones in sequence, depending on the present need. I'll return to this point later on.

1.4.2. Proprioceptive information about ego-movements. Inertial guidance systems measure second order derivatives of position with respect to time, thus they need a twofold integration in order to yield positional information. This is satisfactory as long as the accuracy is high. However, errors are sure to accumulate over time. Such systems are effective in a technical setting (e.g. aircraft, submarines) but appear less suitable solutions in organic systems.

In contradistinction the flow field yields first order derivatives, and if landmarks are used the optical data even yield positional information directly. Thus optic flow structure is a very important data channel in proprioception, especially when maneuvers extend over larger time periods (which explains why e.g. the cruise missile makes use of it for precise navigation near the end of its flight to locate its target or during dodging tactics). That optical information often supersedes vestibular input in the human agent has been convincingly shown several times (Lee and Aronson, 1974). When you try to walk through a long corridor with your eyes closed without hitting the walls the point will certainly impress itself on you painfully.

As has been remarked before it is a simple matter to derive optical information about rotation from the flow. It is also simple to obtain it in the same "format" as the information from the semicircular channels (three orthogonal Cartesian components). Such mechanisms actually exist in vertebrates (Simpson *et al.*, 1981).

This is much less clear with regard to the translational movements, however. It is not hard to devise a method of extracting translation from the flow that could complement the signals from the otolith organs. As has been mentioned before such methods have been used by astronomers for decades. However, a physiologically plausible implementation appears difficult to construct (although most of the calculation implies merely the computation of

weighted sums over space). In some cases the problem is much easier to handle: a prime example is the case of locomotion with respect to a plane (e.g. the floor). The example is a particularly important and interesting one and will be treated in a later section.

1.4.3. Merging and splitting. Since "common fate" is one of the most compelling laws of visual Gestalt (Koehler, 1947), it stands to reason that the mere continuity of optic flow in some region of the visual field leads to an apparent "merging". That region stands out as a single coherent entity, even in the presence of some conflicting evidence from the static image. The merged region need not appear as an object, e.g. in the case of an eye-movement the whole visual field appears as a Gestalt. In this case one that can be discounted as a possible object "out there".

Common fate conflicts, at the boundary of visual objects, have a very strong splitting effect. Even in random noise fields very sharp and vivid apparent boundaries appear the moment two patches move relative to each other (Rogers and Graham, 1979). This is the more striking since in such circumstances the static image is featureless (although textured) in this case.

Mechanisms to detect such boundaries can be of a relatively simple nature. Nakayama and Loomis (1974) proposed some likely candidates.

1.4.4. Extero- and proprioception for the moving observer relative to a plane surface. In this paper I treat two examples at some length, a global problem, namely that of the plane in which the available information is useful for both extero- and proprioception, and a local one, namely that of shape perception in the presence of bending movements. This section treats the former problem.

The case of the moving observer in the presence of a plane surface is a very important practical case and many scientists have worked on the problem. Applications abound in aviation, traffic and ambulatory movements of animals in many natural and artificial environments.

In order to get a feel for the phenomena I offer a few numerical examples. In these examples I show graphically how the flow deforms the visual field. The image of a plane fills at most one half of the optic array, and for reasons of clear exposition it is indeed necessary to show a hemisphere. This can be done by the standard methods of cartography: many projections project the northern hemisphere (or even more)

of the globe on a single flat map. One type of projection is especially convenient because it is *conformal*. This means that a small figure on the optic array is mapped on a similar figure on the plane map: thus the deformations of a gauge figure may be judged especially easily. This type of projection is the stereographic one. In Fig. 5(a) I show a hemisphere of the optic array in stereographic projection, such that the circle bounding the figure coincides with the horizon of the plane. On this projection I have drawn a square checkerboard pattern. Note that this

does not correspond to an even pattern on the (object) plane! There the checks would be deformed just as the checks painted on the object shown in Fig. 2(a). This very fact makes the checks excellently suited as gauge figures.

Now consider the results of some optical flows. Perhaps the simplest case (certainly the one most easily understood intuitively) is that of a movement towards the plane (this happens e.g. when you walk towards a door in a wall). The flow is shown in Fig. 5(b). As was to be expected the checks in the center just blow up

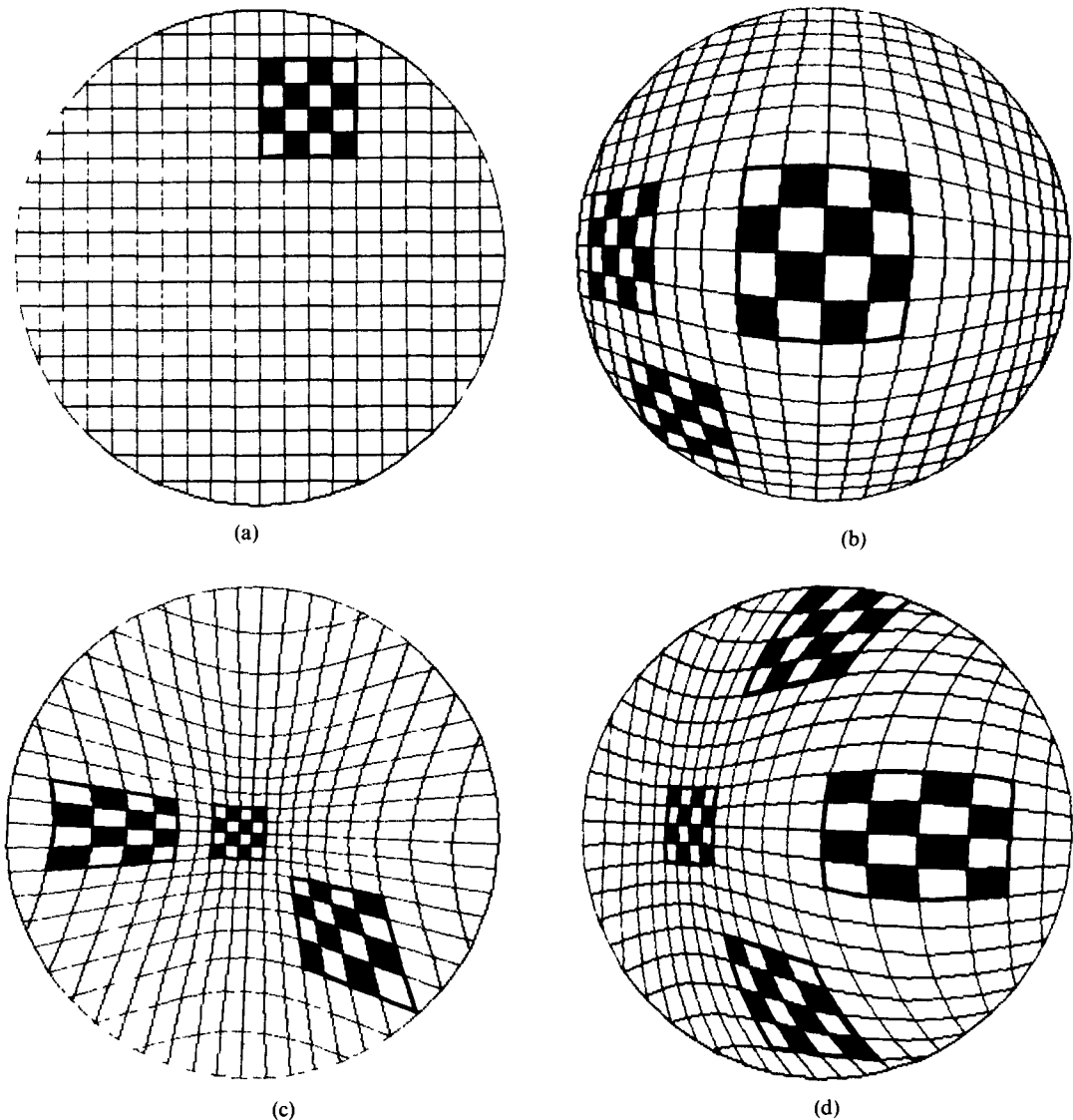


Fig. 5 The optical deformations due to a translation with respect to a plane surface (An added rotation would merely shift or rotate these figures.) The circular circumference is the image of the horizon of the plane. (a) A square grid has been superimposed on the image [This is the same trick as illustrated in Fig. 2(a, b)]. A square gauge figure has been delineated, much like the one used in the sequences illustrated in Figs. 1 and 2. (b) A movement perpendicular towards the plane. (c) A movement in the opposite direction of (b). (d) A translation parallel to the plane, towards the right. In all cases study the deformations of the gauge figure at different positions and compare with the sequences shown in Fig. 2.

(you get closer to them) without any rotation or deformation (after all the nearness gradient vanishes) Perhaps more surprising is the fact that towards the horizon the area of the checks actually shrinks (although you do get closer to them) and that a strong shear appears The shrinkage is due to the fact that you get to view remote parts of the plane more obliquely as you approach the plane These results follow immediately from our formulae, of course. (You should compare the deformations to the examples offered in Figs 1 and 2) Figure 5(c) shows the reverse, a movement away from the plane.

Less immediately intuitively clear is the result of a movement parallel to the plane [Fig. 5(d)] despite the fact that you probably experience such flows on a daily basis (e.g. walking over an even floor) The movement is in the horizontal direction, towards the right. Note that remote parts of the visual field at your right (when walking in the indicated direction) have rotated counter clockwise while patches on the left have rotated in a clockwise fashion. These rotational effects are often spontaneously noticed by people who look out of the (side) window of a train Ahead of you the gauge figure expands in an anisotropic fashion twice as fast in the direction of movement than in the orthogonal direction! The strongest expansion is found at a distance in front of you that exactly equals your eye-height [Thus for car drivers this point will often be obscured by the hood In my (small) car I find it possible to easily study this flow field if I fixate (it takes some exercise to gain proficiency) a point just in front of the car on the highway. Probably an optokinetic nystagmus cancels the average motion and I perceive the local parallax field At an even speed of say 140 km/h I see a flow pattern of the node type, with different expansions in the direction of movement and perpendicular to it. It takes a well textured highway surface to let the experiment succeed.] Behind you the reverse effects occur.

It is especially interesting to note that the deformations as indicated by the fate of the gauge figure yield so very different impressions of the flow than the pattern of field lines does. Thus Gibson (1950, looking only at the pattern of field lines) speaks about the "focus of expansion" which in this case lies on the horizon in the forward direction. Indeed the field lines have their vanishing point there But *the actual expansion vanishes* at the "focus of expansion" as is immediately clear from the gauge figures. The

maximum of expansion is at a location on the floor that is at your eye-height in front of you and not on the horizon at all In the general case (that of a movement in a direction that is oblique with respect to the plane), the extrema of the expansion are located on the bisectrices of the direction of movement and the normal to the plane (Koenderink and van Doorn, 1976a, b, 1978, 1981, Regan and Beverley, 1982).

The extremum of the expansion (the div) is clearly independent of eye-movements, since eye-movements introduce rotations and the div depends only on the translation. On the other hand Gibson's focus can easily be shown to be shifted around during the execution of eye-movements

What is the information available to us from these fields? If the whole field, or at least a large part of it, is available, then you can extract your orientation with respect to the plane, and your instantaneous specific velocity. That is you can find the direction of your movement and the time needed to reach the surface if your instantaneous movement was continued In fact these data can already be obtained from either the div, or the def component alone That the distance is found in temporal units can actually be an asset rather than a drawback in many sensorimotor tasks [e.g. landing reactions (Goodman, 1960, Braitenberg and Taddei-Ferrett, 1966, Lee and Reddish, 1981; Wagner, 1982), avoidance reactions, braking (Lee, 1976), etc.]

1.4.5. *Extraction of shape in the presence of bending deformations* An example of exteroception is that of local shape extraction in the absence of any prior knowledge about eye-movements, etc. I will show that this is possible even in the case that the objects "out there" suffer certain nonrigid transformations

First of all let us consider what is meant by *shape* The narrowest view would be that you know the shape of an object if you can make an exact 3-D copy of it Then you know the three Cartesian coordinates of any point of the object in some position That this definition is uncomfortably narrow is clear from the fact that the human observer may well have an excellent idea of a shape without being able to give dependable estimates of arbitrary Euclidean distances between marks on the object One striking observation is that "visual shape" seems to be invariant against so called "relief transformations" This fact was already noted by Helmholtz (1910) and also by artists For in-

stance, Hildebrand (1913) in his well known technical treatise on the craft and the art of sculpting is very explicit on the issue. Many observers have a hard time to distinguish a relief copy from the same work in the round, except when they are allowed to assume exceedingly oblique viewing positions (technically known as "taking a side view"). A relief transformation conserves collinearity and the intersections of lines. Thus all effects of obscuration are conserved, as are contours, shadow edges and the positions of extrema of light and dark. You know the surface modulo a relief transformation when you know the nearness to the eye for points on the surface except for an unknown additive constant.

In our case the relation is a little different than the one described by Helmholtz, moreover, the "nearness gradient" $\mathbf{F} = \mathbf{grad} \log(d_0/d)$, as introduced earlier, contains yet another ambiguity: it is also degenerate with respect to *scale*. Thus a knowledge of \mathbf{F} (such as may be obtained from the optic flow, *vide infra*) yields the depth modulo a scaling factor and a relief transformation. You may express this in the following way: let d denote the objective distance, d' the inferred one, then d' must be related to d in the following manner:

$$\frac{d_0}{d'} = \alpha \left(\frac{d_0}{d} \right)^\beta$$

d_0 is an arbitrary distance; α, β are arbitrary constants; d_0 and α govern the scale, β governs the relief. Since the vector $\mathbf{F} = \mathbf{grad} \log(d_0/d)$ appears so prominently in the equations it appears likely that monitoring the div, curl and def puts us in a position to find \mathbf{F} and thus the shape (modulo a relief transformation), at least if we are able to solve for \mathbf{F} and \mathbf{A} . It will be shown that this is indeed possible, in fact that this is possible from a knowledge of the def alone!

As a model shape I shall use a polyhedral vertex, and not necessarily a rigid one. You may build such a model from rigid sectors (constant sector angles but they may be mutually unequal) joined with flexible hinges. Thus you obtain flaccid vertices (if there are more than four sectors) that can be flexed. This flexion is a case of bending deformation: distances along the surface remain invariant, although global distances (measured *linea recta*, through the air) in general vary. The model is of especial interest because any real curved surface can be approximated with an arbitrarily fine triangular net and thus by a complicated polyhedron. If you

can solve the "structure from motion problem" for a vertex, then you can treat surfaces on a vertex by vertex basis.

How can we approach the problem? The easiest approach is first to look at the situation at a single edge of the vertex before proceeding to the complete structure.

At an edge you have a discontinuity in the flow field. However, the nature of the singularity is subject to certain constraints: first of all the components of the nearness gradients at both sides of the edge must have identical components along the edge (otherwise the sectors would not hang together at the edge); secondly the same observation can be made for the specific transverse components of the translation (\mathbf{A}_i). This is the case because any relative movement of the facets must be a *rotation around the edge* (remember that the edge functions as a hinge), a property that is conserved in the projection. These two boundary conditions can be combined with the observation of the def at both sides of the edge. The magnitude of the def yields the products of the moduli of \mathbf{A}_i and \mathbf{F} at both sides, whereas the direction of the axis of contraction specifies the bisectrix of \mathbf{A}_i and \mathbf{F} at both sides. These data can be shown to be sufficient for the following computation to be possible: in case \mathbf{A}_{iL} (the vector \mathbf{A}_i on the left of the edge) is specified, the vectors \mathbf{F}_L , \mathbf{F}_R and \mathbf{A}_{iR} can be found. (Actually there are either two possible solutions or none.) Of course one usually has no prior knowledge of \mathbf{A}_{iL} , thus this observation does not at first sight appear to be a step towards any useful solution. But we still have to use the very fact that the spatial structure is a coherent polyhedral vertex.

We can use the fact that the vertex is a coherent surface in the following way. If I make a tour around the vertex over the surface I end up on the same sector at which I started. Thus if I just take any arbitrary choice for \mathbf{A}_i on the first sector, I can do the calculation edge after edge and after some time I end up with a calculated value of \mathbf{A}_i on the same sector at which I started (\mathbf{A}_{i*} say). Now obviously we have the condition $\mathbf{A}_i = \mathbf{A}_{i*}$ to fulfill. This observation is sufficient to solve the problem since the problem is degenerated with respect to scale: we may as well take $\mathbf{A}_i = 1$ (this only changes the arbitrary constant d_0), then we have a single unknown (the direction of \mathbf{A}_i in the visual field) and the equation $\mathbf{A}_i = \mathbf{A}_{i*}$ suffices for a solution. We have tried the scheme in a computer simulation, and it works perfectly. Irrespective of the

precise nature of the rigid movement or the bending [for a N -vertex the bending has $(N - 3)$ degrees of freedom] the algorithm comes up with a unique solution (except for the multiplicity described later on)

Figure 6 shows an example for a hexagonal vertex. (Of special interest because any surface

patch can be approximated by a polyhedron consisting of hexagonal vertices.) Figure 6(a) shows the actual input to the algorithm. This is somewhat simplified even, since you may add arbitrary translations and rotations to the input flow the algorithm does not notice it. The upper two rows in Fig. 6(b) depict what the vertex and

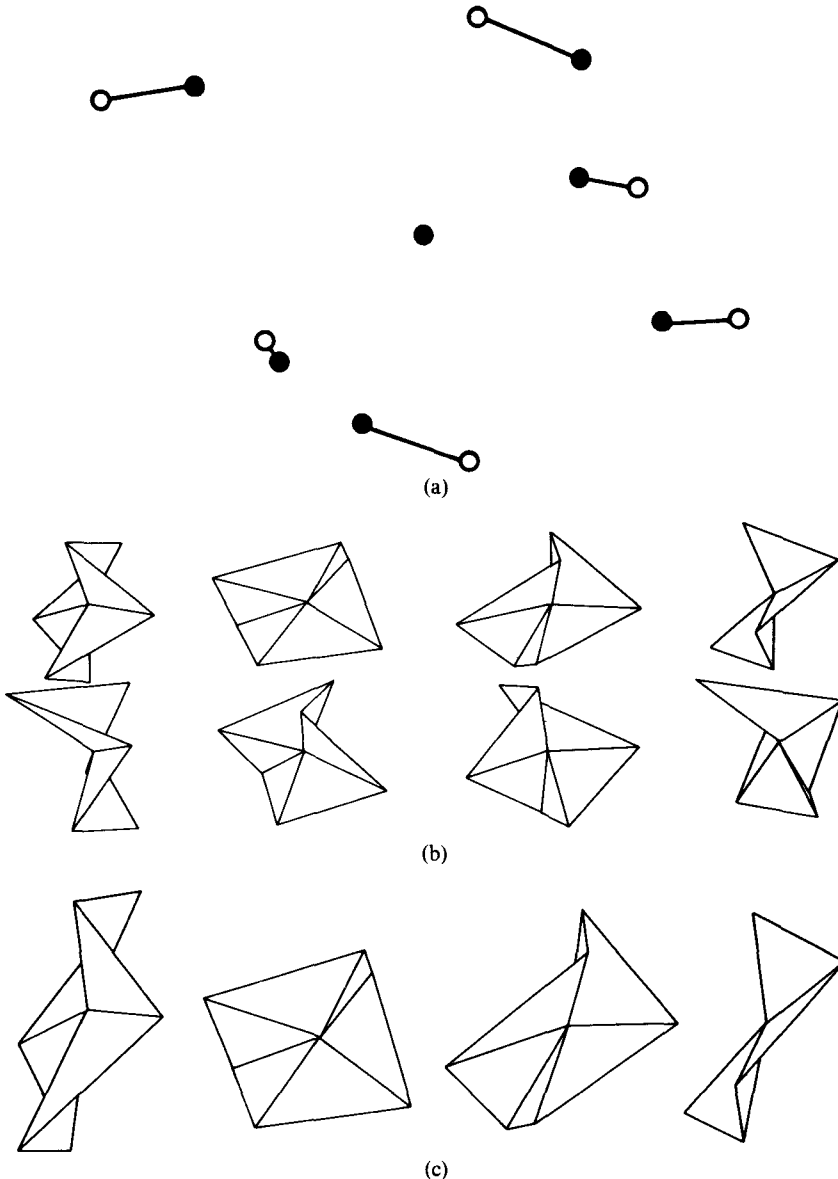


Fig 6 Results of a run of a novel algorithm that yields a partial solution of the "structure from motion problem" in the presence of bending deformations of the object (a) The input to the algorithm the seven defining points of a six vertex in two (temporally close) views, black points the first, open circles the second view The algorithm is completely insensitive to transformations of the second view of the following types —translations, —rotations, —similarities (scale changes) (b) The same configuration as in (a), but seen from quite different vantage points (side views) Edges are drawn for clarity These inputs were not available as input data! The strong bending deformations apparent when you compare the upper row (first view) with the lower one (second view) indicate that there is little use for the "rigidity assumption" here (c) Predictions of the algorithm, on the basis of the data shown in (a), of side views on the moment of the first view [Thus compare (c) with the upper row of (b)!] (In fact the algorithm also extracts the bending itself, not just the shape) Note how well shape can be extracted from this very limited data The final degeneracy (interchange of F 's and A_i 's) was resolved by the algorithm through use of the curl

the vertex after deformation look like from quite different vantage points. These are side views that are not available to the algorithm. The lower row in Fig. 6(c) depicts the predictions based on the result of the algorithm of these same side views. This row should be compared with the upper row of Fig. 6(b). (In fact the algorithm also extracts the deformations.) Note how well the algorithm has caught the shape, although the input conditions violate all requirements of the "structure from motion theorem". (Only two views are given; no four points move together in a rigid fashion) This example shows clearly the value of partial solutions to the problem. (The solution is modulo a relief transformation and a scaling factor.) Many objective features of the shape can be obtained *exactly* even in the presence of these ambiguities, however: this is why it does not seem prudent to despise "merely partial solutions" to the structure from motion problem. After all, organic systems often take shortcuts if useful information can be had cheaply, thus the partial solutions have certainly to be studied as possible paradigms for visual function. Examples of objective properties that can be found even in the presence of scaling and relief ambiguities are: the question of whether a vertex is elliptic or hyperbolic (whether the sum of the sector angles is less or is greater than 360°), or the obscuration relations for side views. (In general all properties depending on the contact of the object with straight lines, thus the very properties that are of importance in reafference concerning the visual system)

The solution is subject to certain ambiguities that should be noted:

- it is degenerate up to a relief transformation;
- it is degenerate up to scale (which obviously any solution has to be),
- it is degenerate as to the sign of A , This means that the solution cannot distinguish between convexities and concavities. Any prior knowledge about movement (e.g. available for ego-movements) disambiguates this situation;
- the solution is degenerate with respect to an interchange between the F 's and A 's.

The latter degeneracy is an especially curious one. It can be solved in several ways. One may e.g. pick the "most rigid" solution. This gener-

ally does the job, but is an *ad hoc* measure. One can also disambiguate the solution by looking at the sign of the curl. This solves the problem. One wonders if the human visual system is subject to illusions caused by this ambiguity. Preliminary psychophysical experiments with deforming vertices on a CRT screen performed by us suggest that this is indeed the case. If this is so, then it would appear that the human does not use curl information for shape extraction, and we have discovered a new, unexpected dynamical perspective illusion.

Note that, except for the possible use of the curl (not for calculation but just to decide between two possible solutions), only the shear (def component) is needed as input data for the solution

2. THE TYPE OF MECHANISMS NEEDED TO EXTRACT INFORMATION FROM THE FLOW: POSSIBLE PHYSIOLOGICAL IMPLEMENTATIONS

I have already commented on the use of mechanisms fit to extract global parameters, like rotation around three perpendicular axes, or the vanishing points due to the translation. I have also noted how detection of mere discontinuity may serve for splitting the visual field into coherent entities. In this section I will comment on mechanisms fit for *local* analysis of flow structure

As remarked earlier there exist many, mutually very different, possibilities to extract local flow structure, e.g. they may be based on relative movement, texture density or orientation and orientation changes of local image detail. First of all let me explain why it appears useful to base the analysis on the differential invariants, no matter how they are extracted

The main virtue of the differential invariants is that they are defined in a coordinate free manner. Their magnitude has a validity that does not depend on the framework in which you measure position, in this respect they are similar to entities like luminance or hue. Div and curl can be extracted locally by mechanisms that are rotationally symmetric. This means that the detection will be receptive field like, concentric structures, in the case of the curl with perhaps a periodic angular dependence. The shear is more complicated since it also involves an orientation, not just a scalar measure. Thus shear detectors must have an orientationally directed structure.

Div detectors have been implicated for the

case of the human observer ("Looming detectors", Regan and Beverley, 1978)

It is feasible that some features of the flow are computed on the basis of (time varying) positional information on image structure. The extremely high acuity reported for some tasks certainly is suggestive of this (Lappin and Fuqua, 1983)

An especially clean way to extract *shear* is to do it by way of the changes in the relative orientations of image detail (Figs 3 and 4). Both the orientation and the magnitude of the shear can be gained from these. Thus *the structures described in the primary visual cortex are exactly the substrate needed for the computation of local shear*. This is especially interesting in view of the fact that monitoring the shear suffices to extract local shape, even in the presence of bending deformations. This seems to indicate a likely function for the system in the primary cortex dedicated to the extraction of local image detail orientation, namely the computation of 3D-shape. (As opposed, or perhaps in addition to, the common notion that this system serves merely the extraction of contour in a two dimensional, rather than a three dimensional setting.)

In summary, the likely mechanisms from a theoretical point of view for which there is at least a trace of evidence in the visual system are.

- mechanisms integrating over large parts of the visual field extracting rotation around one of three mutually perpendicular axes (perhaps coinciding with the normals to the planes of the semicircular channels of the vestibular system) Such systems could aid proprioception and egocentric orientation,
- mechanisms integrating over large parts of the visual field and extracting translation through estimation (in some optimal sense) of the main vanishing points of the flow. Such mechanisms could aid proprioception and the regulation of ego-movement;
- mechanisms integrating over limited parts of the visual field and computing the presence of discontinuities of the flow. Such RF's could form a system for segregation of the visual field,
- mechanisms integrating over limited parts of the visual field and extracting the translation component (or local velocity) of the flow. Such mechanisms ("motion detectors") are certainly present in the visual system of many

vertebrates, often as early as in the retina,

- mechanisms integrating over various regions and coding for expansion (div). Such "looming detectors" have been implied for the human visual system by some psychophysicists. Such systems could extract the immediacy and thus aid in effective locomotion;
- mechanisms integrating over limited parts of the visual field and coding for local shear. Such mechanisms would display strong orientational preference when probed in their subparts. The cortical area 17 is a likely substrate. Such a system would be ideally suited to extract 3D-shape

3. CONCLUSIONS

The study of optic flow has made accelerated progress during recent years, at least in the field of theory. Practical algorithms for use in robotics are in use or forthcoming. However, psychophysical results are extremely scarce as are electrophysiological studies. This is at least in part due to a definite lack of attention to such questions, but also to marked experimental difficulties. Some electrophysiological data are tantalizing in the sense that they appear to indicate the presence of systems dedicated to specific aspects of optic flow analysis. However, because they have generally been interpreted in terms of utility for the analysis of static, two-dimensional images they are usually discussed in a quite different context. Given the fact that theory is so far ahead in this area, there appear to be rich possibilities for the empirical approach here, both in psychophysics and in electrophysiology. In the meantime theoretical developments will continue to speed up because of a practical need from the robotics community.

REFERENCES

- Berthoz A., Pavard B. and Young L. R. (1975) Perception of linear horizontal self-motion induced by peripheral vision (Linearvection). *Expl Brain Res* **23**, 471-489
- Braitenberg V. and Taddes-Ferrett C. (1966) Landing reactions of *Musca domestica*. *Naturwissenschaften* **53**, 155-156
- Braunstein M. L. (1962) The perception of depth through motion. *Psychol Bull* **59**, 422-433
- Cynader M. and Regan D. (1978) Neurons in cat parastriate cortex sensitive to the direction of motion in three-dimensional space. *J Physiol* **274**, 549-569
- Gibson J. J. (1950) *The Perception of the Visual World*. Houghton Mifflin, Boston, Mass.

- Gibson J J, Olum P and Rosenblatt F (1958) Parallax and perspective during aircraft landings *Am J Psychol* **68**, 372-385
- Goodman L J (1960) The landing response of insects I The landing response of the fly *Lucilia sericata* and other celliphorinae *J exp Biol* **37**, 854-878
- Gordon D A (1965) Static and dynamic fields in human space perception *J opt Soc Am* **55**, 1296-1303
- Helmholtz H von (1910) *Treatise on Physiological Optics* (Translated by J P Southall, 1925) Dover, New York
- Hildebrand A (1913) *Das Problem der Form* Heitz & Mundel, Strassburg
- Hoffman D D and Flinchbaugh B E (1982) The interpretation of biological motion *Biol Cybernet* **42**, 195-204
- Koenderink J J and Doorn A J van (1975) Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer *Optica Acta* **22**, 773-791
- Koenderink J J and Doorn A J van (1976a) Local structure of movement parallax of the plane *J opt Soc Am* **66**, 717-723
- Koenderink J J and Doorn A J van (1976b) Visual perception of rigidity of solid shape *J math Biol* **3**, 79-85
- Koenderink J J and Doorn A J van (1978) How an ambulant observer can construct a model of the environment from the geometrical structure of the visual inflow In *Kybernetik 1978* (Edited by Hauske G and Butenandt E) Oldenburg, Muenchen
- Koenderink J J and Doorn A J van (1981) Exterspecific component of the motion parallax field *J opt Soc Am* **71**, 953-957
- Koenderink J J and Doorn A J van (1984) Optical monitoring of ego-motion for movement with respect to a plane surface In *Localization and Orientation in Biology and Engineering* (Edited by Varju and Schnitzler), pp 163-166 Springer, Berlin
- Koehler W (1947) *Gestalt Psychology* Liveright, New York
- Kruk R and Regan D (1983) Visual test results compared with flying performance in telemetry tracked aircraft *Aviat Space Environ Med* **54**, 906-911
- Lappin J S and Fuqua M A (1983) Accurate visual measurement of three-dimensional moving patterns *Science* **221**, 480-482
- Lee D N and Aronson E (1974) Visual proprioceptive control of standing in human infants *Percept Psychophys* **15**, 529-532
- Lee D N (1976) A theory of visual control of braking based on information about time-to-collision *Perception* **5**, 437-457
- Lee D N (1980) The optic flow field the foundation of vision *Phil Trans R Soc Lond B* **290**, 169-179
- Lee D N and Reddish P E (1981) Plummeting gannets a paradigm of ecological optics *Nature* **293**, 293-294
- Longuet-Higgins H C and Prazdny K (1980) The interpretation of moving retinal images *Proc R Soc Lond B* **208**, 385-387
- Longuet-Higgins H C (1981) A computer algorithm for reconstructing a scene from two projections *Nature* **293**, 133-135
- Mach E (1962) *The Analysis of Sensations* (English translation 1962, originally published in 1886) Dover, New York
- Metzger W (1934) Tiefenerscheinungen in optischen Bewegungsfeldern *Psychol Forsch* **20**, 195-206
- Nakayama K and Loomis J M (1974) Optical velocity patterns, velocity-sensitive neurons, and space perception a hypothesis *Perception* **3**, 63-80
- Prazdny K (1980) Egomotion and relative depth-map from optical flow *Biol Cybernet* **36**, 87-102
- Regan D and Beverley K I (1978) Looming detectors in the human visual pathway *Vision Res* **18**, 415-421
- Regan D, Beverley K and Cynader M (1979) The visual perception of motion in depth *Scient Am* **241**, 136-151
- Regan D and Beverley K I (1979) Visually guided locomotion psychophysical evidence for a neural mechanism sensitive to flow patterns *Science* **205**, 311-313
- Regan D and Beverley K I (1982) How do we avoid confounding the direction we are looking and the direction we are moving? *Science* **215**, 194-196
- Regan D and Beverley K I (1983) Psychophysics of visual flow patterns and motion in depth In *Festschrift fur Professor I Kohler* (Edited by Spillmann and Wooten) Lawrence Erlbaum, Hillsdale, N J
- Rogers B and Graham M (1979) Motion parallax as an independent cue for depth perception *Perception* **8**, 125-134
- Rosenberg K (1924) Ueber eine Gesichtstauschung *Z Physik Chem Unterr* **37**, 160-164
- Simpson J I, Graf W and Leonard C (1981) The coordinate system of visual climbing fibers to the flocculus In *Progress in Oculomotor Research* (Edited by Fuchs A F and Becker W), pp 475-484 Elsevier, New York
- Stevens K A (1981) The information content of texture gradients *Biol Cybernet* **42**, 95-105
- Stevens K A (1983a) Surface tilt (the direction of slant) a neglected psychophysical variable *Percept Psychophys* **33**, 241-250
- Stevens K A (1983b) Slant-tilt the visual encoding of surface orientation *Biol Cybernet* **46**, 183-195
- Todd J. T (1982) Visual information about rigid and non-rigid motion a geometric analysis *J exp Psychol* **8**, 238-252
- Ullman S (1979) The interpretation of structure from motion *Proc R Soc Lond B* **203**, 405-426
- Ullman S (1980) Against direct perception *Behav Brain Sci* **3**, 373-415
- Wagner H (1982) Flow-field variables trigger landing in flies *Nature* **297**, 147-148
- Whittaker E T (1944) *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (First printing 1904) Dover, New York
- Whiteside T C D and Samuel G D (1970) Blur zone *Nature* **225**, 94-95