Families of stationary patterns producing illusory movement: insights into the visual system

CORNELIA FERMÜLLER, ROBERT PLESS AND YIANNIS ALOIMONOS

Computer Vision Laboratory, Center for Automation Research, Institute for Advanced Computer Studies, Computer Science Department, University of Maryland, College Park, MD 20742-3275, USA

SUMMARY

A computational explanation of the illusory movement experienced upon extended viewing of *Enigma*, a static figure painted by Leviant, is presented. The explanation relies on a model for the interpretation of three-dimensional motion information contained in retinal motion measurements. This model shows that the *Enigma* figure is a special case of a larger class of figures exhibiting the same illusory movement and these figures are introduced here. Our explanation suggests that eye movements and/or accommodation changes cause weak retinal motion signals, which are interpreted by higher-level processes in a way that gives rise to these illusions, and proposes a number of new experiments to unravel the functional structure of the motion pathway.

1. INTRODUCTION

A visual illusion has recently given rise to a controversy in the vision research community. This illusion arises when viewing *Enigma*, a static figure painted by Leviant. *Enigma*, shown in figure 1*a*, consists of radial lines emanating from the centre of the image and interrupted by a set of concentric uniformly coloured rings. Upon extended viewing of *Enigma*, most human subjects perceive illusory movement inside the rings which keeps changing direction.

The nature of the controversy lies in the view of researchers regarding the location in the brain where the illusion takes place and the involved processes. Zeki (1994, 1995) and his co-workers (Zeki *et al.* 1993) argue that higher-level processes are responsible for the perception of illusory movement. Thus, in their view, the *Enigma* static stimulus 'induces activity in a given region of the visual cortex which then invests the stimulus with a particular perceptual quality, the latter being entirely the construct of the brain'. Their arguments are based on measurements of changes in regional cerebral blood flow (rCBF) using the technique of positron emission tomography. When comparing the rCBF measurements during the viewing of *Enigma* and the reference image shown in figure 1*b* (radial lines pass through the rings), they found comparable values in V1—an early area in visual processing, and for *Enigma* significantly higher values in V5 (or MT)—a later area in visual motion processing. In a diametrically opposed viewpoint, Gregory (1993, 1995) argues for low-level processes as the cause of the illusion. He suggests that the same processes responsible for the MacKay illusion (see figure 1*c* and MacKay (1957, 1958)) cause illusory movement in *Enigma*. He proposes that 'the dynamic shimmer seen in these static figures has an optical cause: changes of size of the retinal image due to the usual rapid 'hunting' of accommodation'. A third opinion was recently published (Mon-Williams & Wann 1996) suggesting that small roto-translational eye movements may be the cause of the illusion.

In this paper we present a computational explanation of the Leviant illusion. Our view is that the perception of illusory movement is due to the particular architecture of the visual motion pathway, which takes as input, weak retinal motion measurements due to small eye movements. We agree with Gregory that weak retinal motion signals due to accommodation or eye movements play an important role in the creation of the illusion, but while such retinal motion signals are sufficient to explain the MacKay effect, they cannot by themselves explain the Leviant illusory movement. We also agree with Zeki that the reason for the illusion lies mainly in high-level processes. These processes must be triggered by retinal motion signals, however weak they are. The major contribution of this work lies in the explanation of these higher-level processes in computational terms.

The basis of our approach is the computational theory of visual motion interpretation that we have developed. Since autonomous systems, biological or artificial, move in their dynamic and changing environments, the most basic tasks they need to perform are the understanding of their own three-dimensional motion as well as those of other moving entities. Thus, they have to solve the so-called egomotion estimation problem and the motion segmentation problem. We have developed a theory that accounts for the robust and efficient estimation of three-dimensional motion (see Fermüller 1995; Fermüller & Aloimonos 1995*a*, *b*, 1997). This theory postulates a general architecture for its implementation, and

this architecture is consistent with neurobiological findings concerning the motion pathway in primates. Our hypothesis is that this architecture can serve as a working model for the human motion pathway and it is on this basis that we can explain the Leviant illusion.

The gist of the computational theory, as explained in the next section, is that retinal motion measurements along various orientations all over the image are grouped in particular ways. Each group consists of motion measurements along predefined orientation fields, which are called coaxis and copoint fields and possess a number of important properties making three-dimensional motion estimation most easy from a computational perspective.

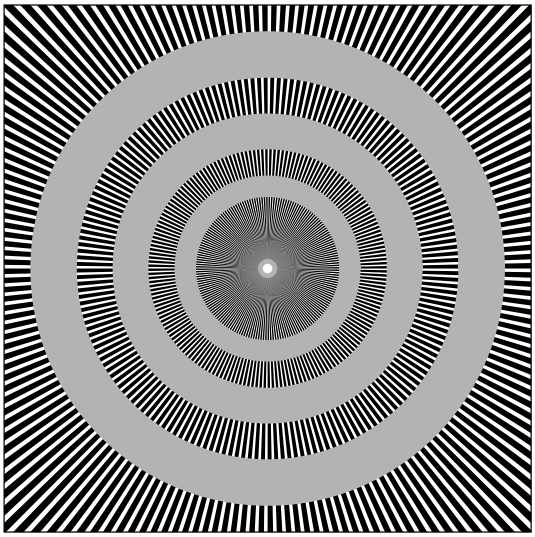
The *Enigma* figure has the property that all its markings can give rise to motion vectors belonging to only one of the predefined orientation fields. If indeed the Leviant illusion is an artifact of the particular way the motion pathway is constructed for the purpose of three-dimensional motion estimation, then any other image in the style of *Enigma*, with the property that all its markings can give rise to motion vectors belonging to only one of the predefined orientation fields, must give rise to a similar illusory movement. We show that this is indeed the case. On the basis of our computational theory we can produce a large number of images that give rise to illusory movement, some of which will be shown here. The theory also predicts that other figures not based on the predefined orientation fields, even if they are superficially similar, will not cause the same illusory movement.

The organization of the paper is as follows. Section 2 explains the computational model for the interpretation of visual motion on which this paper is based. Section 3 provides a number of new images giving rise to illusory movement, § 4 demonstrates that the illusion cannot be due to a MacKay effect and gives a detailed explanation of the computational steps leading to the experience of it, and § 5 concludes the work.

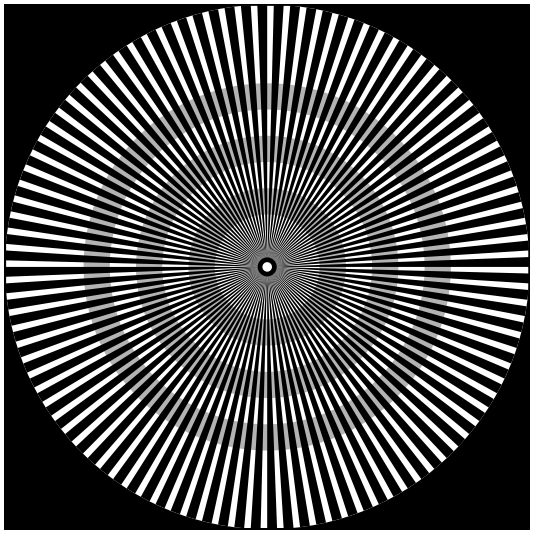
2. A COMPUTATIONAL THEORY OF THE ANALYSIS OF VISUAL MOTION

Any system that moves in its environment has to reach an understanding of its own motion. Although an organism may move in a non-rigid manner as a whole, with its head, arms, legs or wings undergoing different motions, the eyes move rigidly—that is, as a sum of an instantaneous translation and rotation. Thus, any system has to interpret the rigid motion encoded in the sequence of images perceived by its retinae.

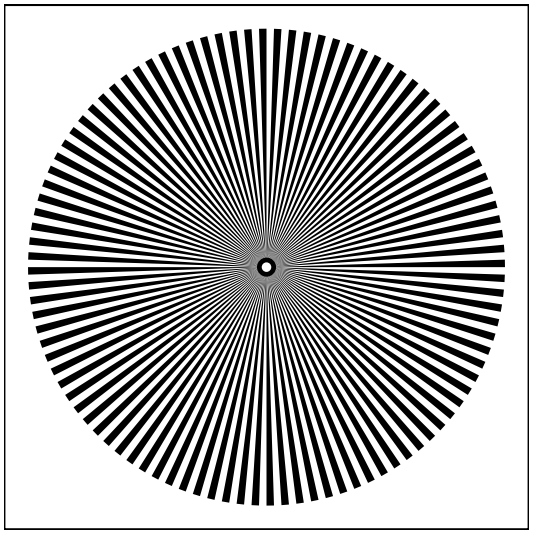
As a system moves in its environment, each point of the environment has a three-dimensional velocity vector with respect to the system. To describe the perceived motion on the image plane the concept of the visual motion field has been used. This field corresponds to the projection of the three-dimensional velocity vectors on the image plane; it depends on



(*a*)



(*b*)



(*c*)

Figure 1. (*a*) *Enigma*, giving rise to the Leviant illusion: fixation at the centre results in perception of a rotary motion inside the rings. (*b*) The control figure used in Zeki *et al.* (1993) in which there is no perception of rotary motion. (*c*) MacKay rays, giving rise to a shimmering at the centre and a circular propeller-like movement in the other parts of the figure.

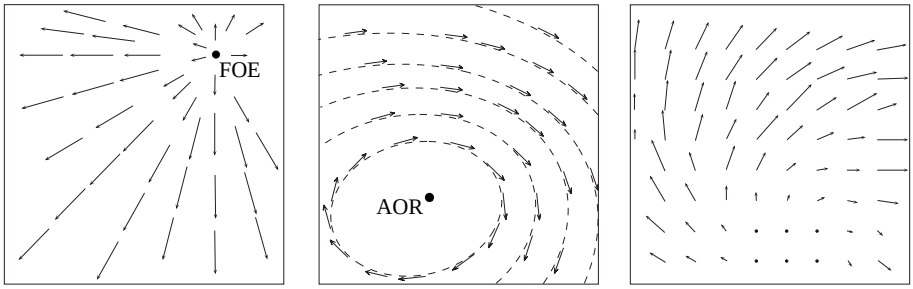


Figure 2. (*a*) Purely translational; (*b*) purely rotational; and (*c*) general motion field due to rigid motion.

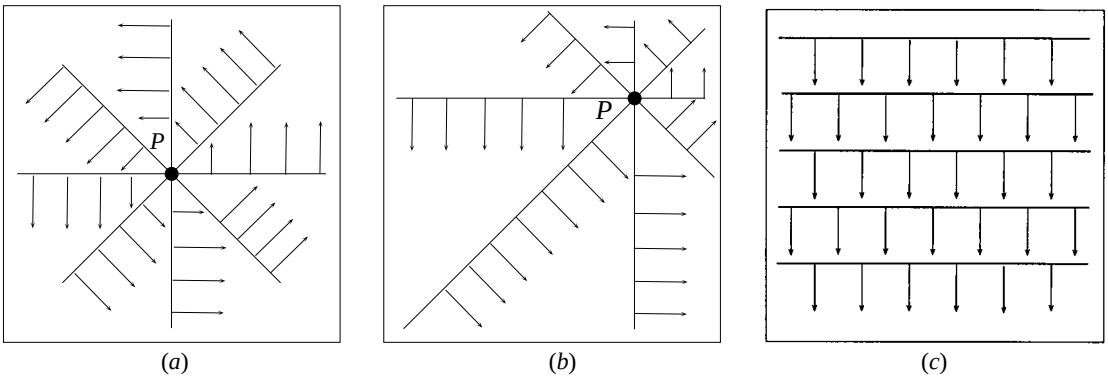


Figure 3. Copoint fields.

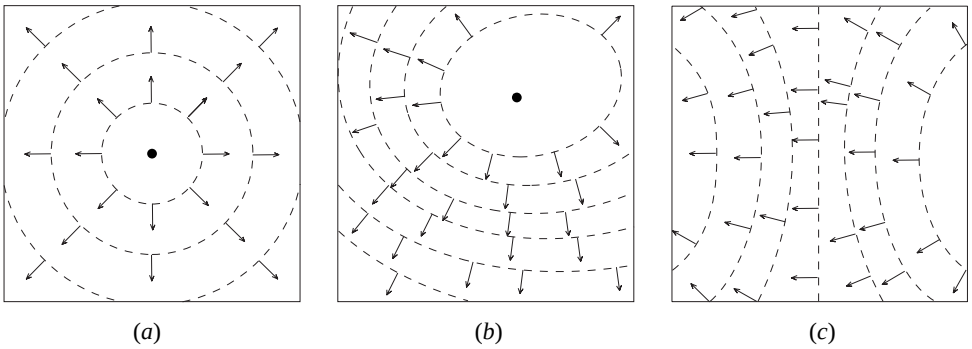
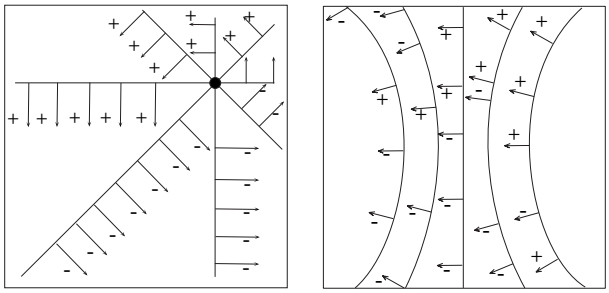


Figure 4. Coaxis fields.

the three-dimensional motion and the structure of the scene in view. If the system's motion is purely translational the motion has the structure of the field shown in figure 2*a*. It consists of vectors emanating from the point where the translation axis pierces the image plane, called the focus of expansion (FOE) or focus of contraction (FOC), depending on whether the observer approaches or moves away from the scene. The length of each vector is inversely proportional to the distance from the scene. If the system's motion is purely rotational the motion field has the structure of the field shown in figure 2*b*. It consists of vectors tangential to conic sections defined by the intersection of the image plane with the family of cones having the rotation axis as their axis. The point of intersection of the rotation axis with the image plane is called the axis of rotation (AOR). If the system moves in a general manner involving both translation and rotation then the motion field amounts to the addition of a translational and rotational field and no longer has a simple structure (for an example, see figure 2*c*). The problem of estimating a system's three-dimensional motion then becomes equivalent to the hard problem of decoupling the translation from the rotation and recognizing the positions of the FOE and AOR.

In actuality, the problem of three-dimensional motion estimation is even harder because the measurements of retinal motion that can be made on the image are due to movements of light patterns only, and this is not equivalent to the motion field. To denote the exact movement of every point in the image, the term optical flow field has been used. Accurate values of the optical flow field, however, cannot possibly be computed; in general, on the basis of local information, only the component of the flow perpendicular to edges in the image can be estimated—this is the well-known aperture problem. In many cases, it is possible to obtain additional flow information (e.g. from features of high curvature, corners). Thus, the



(*a*) (*b*)

Figure 5. Copoint and coaxis fields filled with qualitative motion measurements.

input used for further motion processing is some *partial optical flow* information. A system, in order to reliably estimate its three-dimensional motion, must make use of retinal motion measurements globally and in a redundant way. We have shown that this is possible by using only the signs of flow measurements in various directions. The major idea is based on selecting groups of measurements of the sign of retinal motion along predefined directions, which form patterns in the image. These patterns encode the three-dimensional motion in a simple manner. The rest of this section summarizes the theory. For more information, the reader is referred to Fermüller (1995) and Fermüller & Aloimonos (1995*a*, *b*, 1997).

We will consider two classes of orientation fields that are essential in our approach. Along these orientation fields, image motion measurements will be grouped together.

The first class, called copoint fields, is defined by one point, P, on the image plane (this point could lie inside or outside the image). A copoint field consists of unit vectors that at every image point are perpendicular to the line connecting that point with P. Figure 3*a*–*c* shows three examples of copoint fields with point P: at the centre of the image (*a*); in the periphery of the image (*b*); and at infinity (*c*).

The second class, called coaxis fields, is defined by one axis (that could intersect or be parallel to the image plane) passing through the origin—nodal point—of the eye. Considering one axis, $\boldsymbol{s}$, the coaxis field is defined by the unit vectors that are normal to the conic sections resulting from the intersection of the image plane with the family of cones having $\boldsymbol{s}$ as their axis. Figure 4*a*–*c* shows three examples of coaxis fields. In (*a*), axis $\boldsymbol{s}$ is perpendicular to the image at its centre. In (*b*), axis $\boldsymbol{s}$ intersects the image at point s, and in (*c*) axis $\boldsymbol{s}$ is horizontal and parallel to the image plane.

Our method proceeds by making measurements of the sign of retinal motion along the orientations of one coaxis or copoint field, and it does so for several fields. When measurement of image motion is performed at a point (x, y) along a direction $\boldsymbol{n}$, there are three possibilities. Either image motion is measured in the same direction as $\boldsymbol{n}$, in which case we denote the measurement by '+'; image motion is measured in the opposite direction, where the measurement is denoted by '−'; or there is no motion, and the measurement is denoted by '0'. Such measurements can be easily made using Reichardt-like detectors or equivalent energy models (Adelson & Bergen 1985; Grzywacz *et al.* 1995; Reichardt 1961, 1987; Snippe & Koenderink 1994; Van Santen & Sperling 1984).

Thus, the input to the system is a number of coaxis and copoint fields with every vector in each field labelled '+', '−' or '0', and it is on the basis of these 'maps' that subsequent processing for three-dimensional motion estimation will take place. (Figure 5 shows one copoint and one coaxis field 'filled' with 'qualitative' motion measurements.)

The maps, consisting of the discrete measurements +, − and 0, encode the three-dimensional motion in the form of global patterns. We provide next a simple example of such a pattern and we refer later to the general case.

Consider the coaxis vectors due to an axis perpendicular to the image at its origin. (As shown in figure 5*a*, these vectors are perpendicular to the circles centred at the origin.)

Assume first that there is only translation. In figure 6*a*, which illustrates this situation, the coaxis vectors are shown in boldface and the motion vectors (which are actually unknown) in grey. Clearly, the motion vectors emanate from the FOE. The value of the sign of retinal motion along a coaxis vector at a point depends on the angle this vector makes with the motion vector (since measurement of retinal motion along a coaxis vector amounts to the projection of the motion vector at that point on the coaxis vector).

For the coaxis vectors at the points lying on the circle with diameter equal to the segment connecting the FOE to the image centre, the angle between the motion vector and the coaxis vector is 90°. For the points inside the circle the angle is greater than 90° and for the points outside it is smaller than 90°. This means simply that inside the circle the measurement of the sign of retinal motion along the coaxis vectors will be negative, outside the circle positive, and on the circle zero.

If there is only rotation, a similar pattern is obtained. Rotation exists around all three axes, x, y and z, but the rotation around the z-axis does not contribute any value to the image motion measurements along these coaxis vectors because the flow it produces is perpendicular to them. In this case, measurements of the sign of image motion along the coaxis vectors will be positive and negative, separated by a straight line passing through the image origin. The line represents the projection of the rotation axis (α, β, γ) on the image plane (figure 6*b*).

For the general case of rigid motion we need to combine the results of figure 6*a*, *b*. Superimposing the two figures we find that whenever positive and positive come together the result will be positive, while when negative meets negative the result will be negative. Whenever positive and negative come together the result depends on the actual values and cannot be known without knowledge of the depth of the scene.

Thus, for the coaxis vectors considered, the measurements of the sign of image motion along them

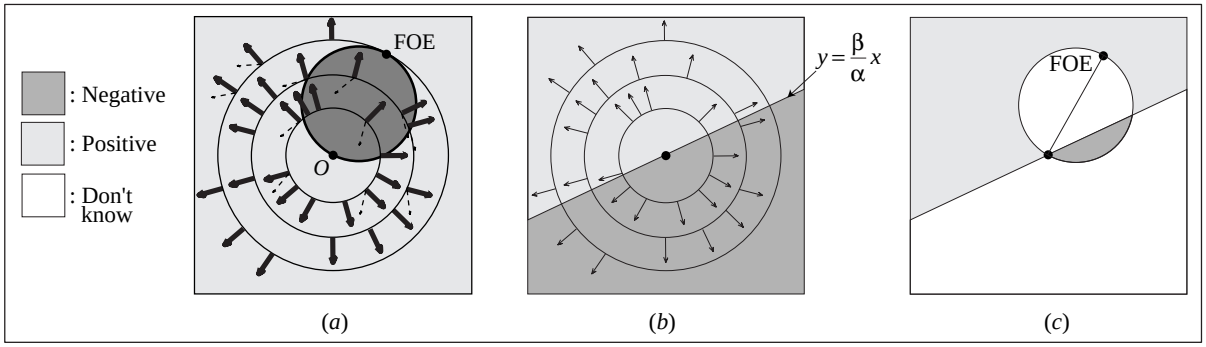


Figure 6. Motion pattern for a coaxis vector with centre at the origin. (*a*) Motion measurements due to translation are separated by a circle, which passes through the FOE, according to the signs of their values $(+, -, 0)$. (*b*) Separation of motion measurements due to rotation by a line passing through the AOR, lying somewhere on the line $y = (\beta/\alpha)x$. (*c*) Motion pattern for a general rigid motion.

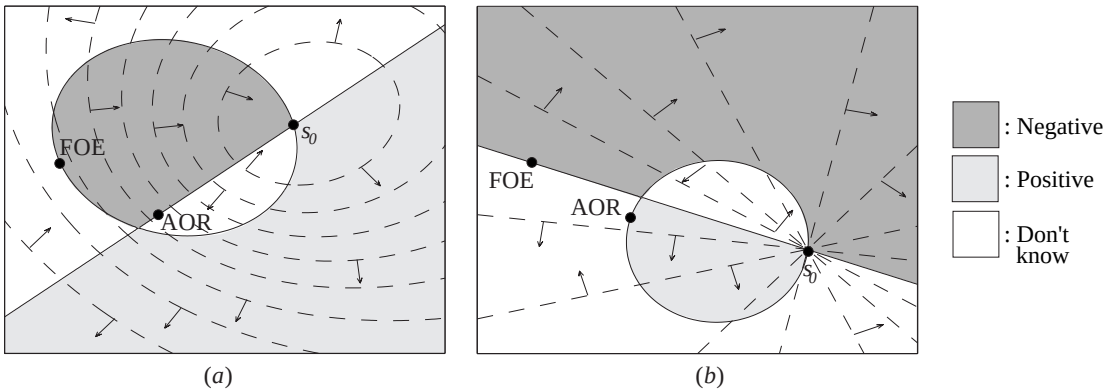


Figure 7. (*a*) Coaxis vectors and (*b*) copoint vectors in the plane with the patterns superimposed. Note that the 'don't know' area contains both positive and negative vectors. In (*a*) s_0 is the point where the axis pierces the image plane and in (*b*) s_0 is the point defining the field.

form a global pattern on the image defined by a circle and a straight line (figure 6*c*), defining an area of only positive values, an area of only negative values and an area where both positive and negative values are possible depending on the scene. Localization of this global pattern in the image provides information about three-dimensional motion. Indeed, the FOE is the antidiametric point of the origin, and the line represents the projection of the rotation axis on the image.

Figure 7 describes the global patterns for general coaxis and copoint fields. In both cases the pattern is defined by a straight line and a conic section. For coaxis fields the conic section separates the translational measurements into positive and negative, and the straight line separates the rotational measurements. For the copoint fields we have the dual situation. In figure 7*a*, the conic section passes through the FOE and the straight line through the AOR. In figure 7*b*, the dual situation is observed.

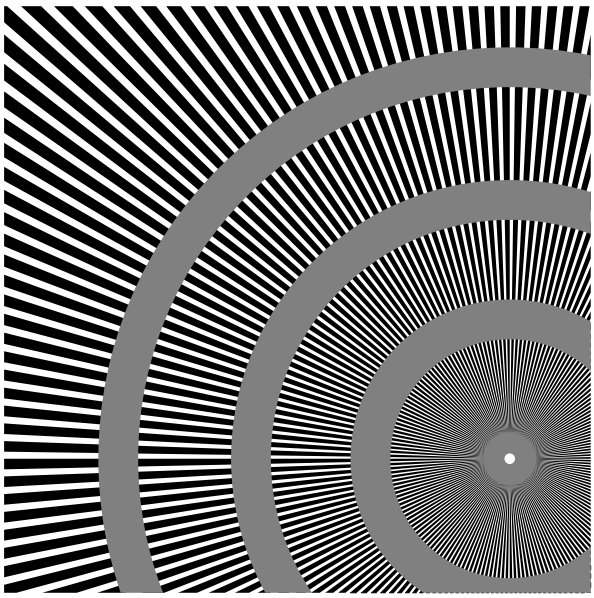
Thus, performing measurements along the vectors of several coaxis and copoint fields, we obtain many patterns which result in an unambiguous determination of the FOE and the AOR (see Fermüller (1995) and Fermüller & Aloimonos (1995*a*, *b*, 1997) for details).

These patterns are simple in the sense that they reduce the complexity of the three-dimensional motion estimation problem in the most efficient way possible. While general rigid motion is encoded in the image plane in the form of five parameters—two for the direction of translation and three for the rotation—each of these depends on only three parameters. The coaxis and copoint fields underlying these patterns are the only vector fields that can give rise to such a simplification.

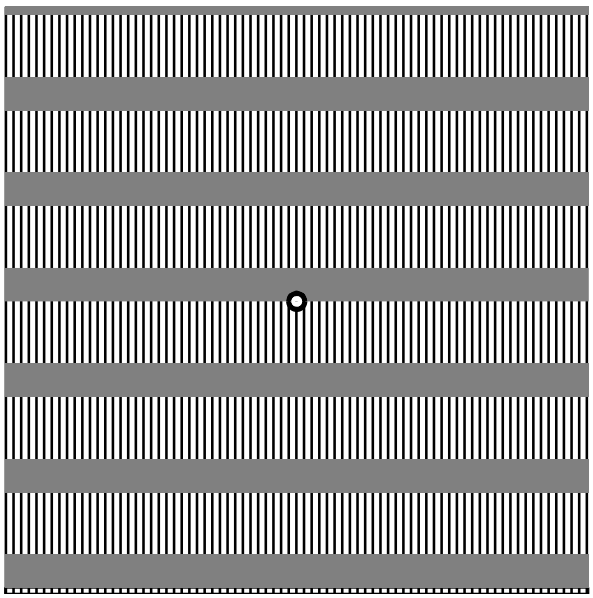
The theory described above can be used by any system with sufficient computing power to estimate the FOE and AOR very robustly from the signs of image velocity measurements. This is not to say that all aspects of motion estimation must be based on this theory. Undoubtedly, any system uses additional extraretinal information to solve this problem, such as inertial sensor information (Aloimonos 1993). In addition, systems employ estimates of the values of the image velocity measurements in different ways to obtain velocity information.

3. NEW ILLUSIONS IN THE STYLE OF *ENIGMA*

There exists a relationship between the structure of the *Enigma* figure and the copoint vector field with centre, P, at the centre of the image. The radial rays of *Enigma* are perpendicular to these copoint vectors and the homogeneous rings are tangential to them,



(*a*)

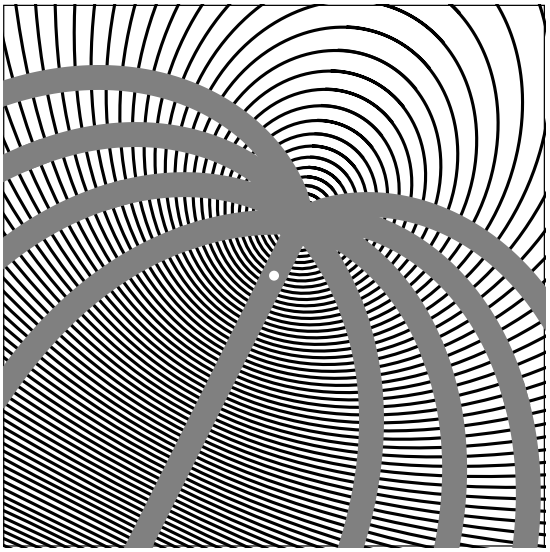


(*b*)

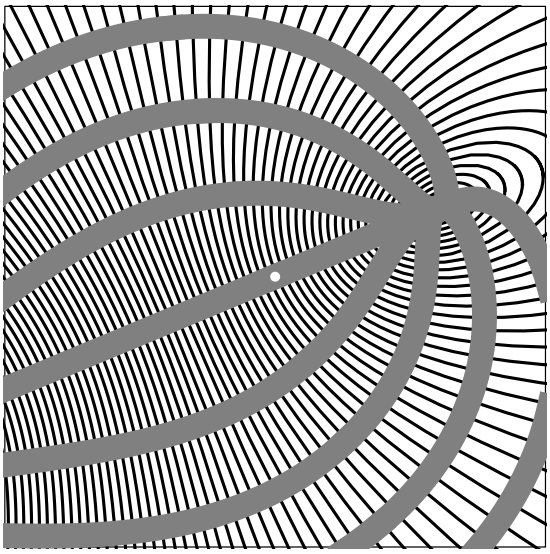
Figure 8. Illusory movement figures based on copoint fields. Experience of the illusion is facilitated by fixating on the white dot displayed at the image centre.

and thus normal to the rays. We applied this relationship to obtain new figures from different copoint fields. Figure 8 shows two such examples with the centre of the copoint field in 8*a* being at the periphery of the image, and in 8*b* at infinity. Both figures produce illusory movement inside the bands. It is worth noting that example 8*b* is well known in the literature (MacKay 1961).

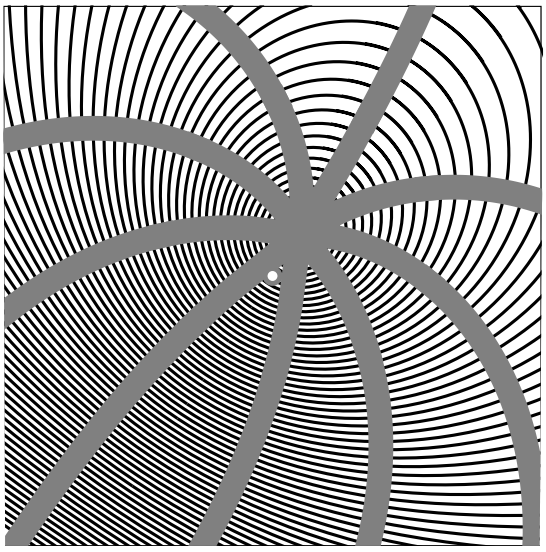
In a similar manner we can apply the same principle to coaxis vector fields and obtain new figures exhibiting the same illusory movement. In particular, given a coaxis vector field, we choose rays perpendicular to the coaxis vectors and homogeneous bands tangential to them. Figure 9*a–c* gives examples of three different coaxis fields.



(*a*)

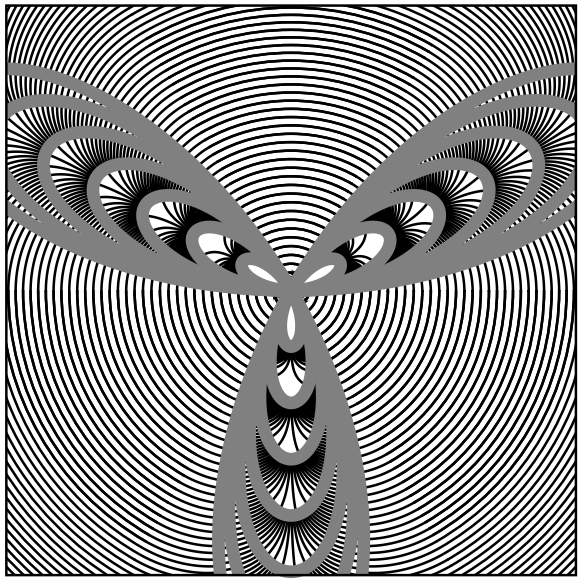


(*b*)

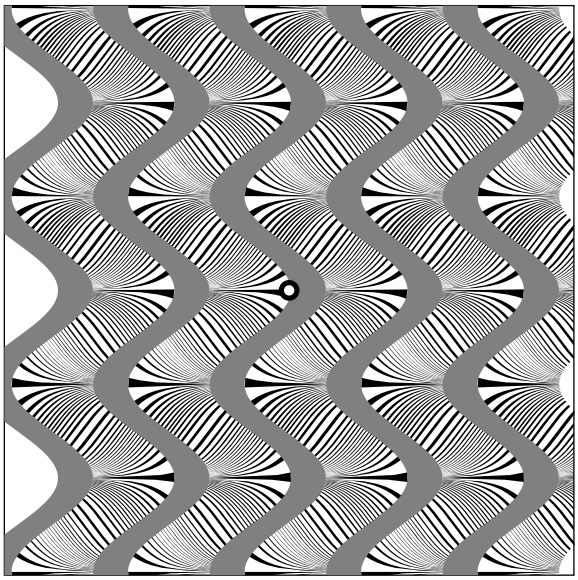


(*c*)

Figure 9. Illusory movement figures based on coaxis fields. As the coaxis vectors are defined by the intersection of the image plane with a family of cones, the related figures depend on the focal length. Here the focal length is chosen to be about equal to the size of the image. Thus, the illusions are best experienced at a short distance.



(*a*)



(*b*)

Figure 10. No significant illusory movement is perceived with figures based on vector fields different from coaxis and copoint fields.

No significant illusory movement is experienced when the figure is constructed by applying the same principle to vector fields different from coaxis or copoint fields. This is demonstrated in figure 10, showing 'leaf' and 'sinusoidal' patterns, respectively.

4. EXPLANATION OF *ENIGMA* AND RELATED ILLUSIONS

A look at motion interpretation from a purely computational point of view gives us insight into the complexity and possible structure of the computations involved in three-dimensional motion estimation and segmentation. In § 2 a theory was described that showed how to estimate the three-dimensional motion of a system moving rigidly in a static environment on the basis of global patterns of the spatiotemporal measurements. Turning from the theory to an implementation, we have to consider the fact that systems moving in the real world don't move in static environments. They are confronted with scenes containing objects moving in various ways. In order to estimate three-dimensional motion from patterns, the image must be segmented into areas of coherent motion; however, the segmentation problem for a moving system is very difficult. It is an ill-posed problem and at this point we do not know the best way to solve it. We do, however, understand some properties that an optimal solution must have.

Segmentation involves both detection and localization, the two problems giving rise to the antagonistic conflict at the heart of pattern recognition. From the viewpoint of signal processing, it has been proved (Gabor 1946) that any single (linear) operator can answer only one of these questions with sufficient accuracy. In order to detect or recognize the existence of different motions within image patches, a sufficiently large number of flow measurements must be available. To obtain the measurements, some form of smoothing or averaging in the spatial as well as the temporal domain has to be done. On the other hand, the localization of objects requires the use of very accurate local flow information.

It can easily be understood that it is not possible to perform the segmentation solely on the basis of image measurements. The reason is that motion, as well as depth discontinuities, manifest themselves as discontinuities in the flow field. Furthermore, even if different motions could be separated, there remains the question of which motion belongs to the static background and which motions belong to the independently moving objects. In order to segment successfully, some additional three-dimensional motion information and probably some shape information about the scene accumulated over time has to be taken into account.

The logical consequence of the interplay between the two problems of motion estimation and segmentation is that they are closely coupled and must be solved simultaneously. An optimal solution for segmentation requires the system to first detect different motions in the image and obtain some estimate of the three-dimensional motion with the use of a global flow field representation. The derived estimates then have to be utilized to accurately localize and also recognize independently moving objects from local flow field information. In an additional step the results of the segmentation could be used for more accurate motion estimation.

From the biological side, in the past decades much information has been uncovered about the structure and simple functional properties of neurones in the visual motion pathways of primates (Zeki 1993). It is well established that from the earliest stages of the Pa ganglion cells and magnocellular LGN cells, over the various layers in V1 to area MT and MST, an increase in the sizes of the receptive fields occurs. Furthermore, an onset of orientation selectivity has been observed, when a bar is moved perpendicular

to its direction. Whereas the neurones in the earliest stages (Pa ganglion cells, magnocellular LGN cells and cells of layer 4Ca in area V1) respond almost equally well to every direction, cells in layer 4C of V1 show a preference for a particular orientation and in layer 4B and 6 of V1 as well as in V2 and MT cells are direction-selective (Movshon 1990) (see figure 11). For cells later in the motion pathway in MT and MST, which can have very large receptive fields ranging from 30 to 100 °C, selectivity to particular three-dimensional-motion configurations has been reported (Andersen *et al.* 1990; Duffy & Wurtz 1991*a*, *b*; Lagae 1991; Orban *et al.* 1992; Tanaka & Saito 1989).

The neurobiological findings just described do not contradict our theory of motion estimation. One can easily envisage an architecture that, using neurones with the properties of those in a primate's motion pathway, implements a global decomposition of the motion field using the signs of motion vectors along appropriately chosen directions. Neurones of the kind shown in figure 11*a* could be involved in the estimation of local retinal motion information; they could be thought of as computing whether the projection of retinal motion along some direction is positive or negative. Neurones of the kind shown in figure 11*b* could be involved in the selection of local vectors in particular directions, as parts of the various patterns discussed. Neurones of the kind shown in figure 11*c* could be involved in computing the signs (positive or negative) of pattern vectors for areas in the image; that is, they might compute, for image patches of different sizes, whether the flow there is positive or negative. Finally, neurones of the kind found in MT and MST could be the ones that piece together the parts of the patterns previously found, into global patterns that are matched with prestored global patterns. Such matches provide information about three-dimensional motion.

The knowledge we have about the neurobiology of the motion pathway, together with the computational theory of motion estimation and the computational arguments regarding segmentation, provide us with an explanation for the illusions described above.

Before we give further details we apply our computational principles to show that the illusory movement seen in the MacKay rays can be explained by processes taking place at very early stages. Subsequently, we discuss why the *Enigma* figure, and our new figures, defy this type of explanation.

MacKay reported illusions of movement in images of repetitive patterns. For rays emanating from the image centre (see figure 1*c*) he describes (MacKay 1958) two different kinds of movement: a shimmer in the middle of the image and a circular movement away from the centre.

Small eye movements due to instability as well as small accommodative movements give rise to retinal motion signals. For simplicity we model this movement as a translation along the x-axis, but similar results would follow from other movements. At the very early processing stages the receptive fields of cells are very small, and thus the motion measurements generated, e.g. by biological models like Reichardt detectors or by energy-based models, approximate normal flow measurements (i.e. motion measurements perpendicular to edges). For a small translational motion along the x-axis, the normal flow field is shown in figure 12. The irregularity of the motion in the centre of the figure—the shimmering—is created because the translational image motion is greater than the separation of the rays, causing the local motion detectors to give results that are incorrect and vary rapidly as functions of their spatial coordinates (the so-called aliasing problem). Outside of this central region, the motion field is both spatially smooth and consistent with motion along the complementary patterns proposed by MacKay (1957). A random sequence of small translational motions of the eye will contain subsequences of temporally consistent image motion, giving rise to the occasional rapid propeller-like motions around the edge of the figure.

This explanation, involving only low-level processes, cannot possibly account for the movement seen in the Leviant illusion.

If we consider a small translational eye movement, we obtain a motion field like the one in figure 13*a*. Since the regions inside the rings are completely homogeneous and there are no features to provide input to the motion detectors, no motion is detected there. Regardless of accommodation movements, microsaccades or slow drifts of the eye, the putative causes of the MacKay motion (Gregory 1993; Mon-Williams & Wann 1996), the retinal input on the region of the retina that is seeing the rings does not change. However, the clear perception is of a motion exclusively within the rings (figure 13*b*).

One might ask whether processing very early in the motion pathway (if not the retinal signals themselves) could create the perceived movement. Such processing could only consist of spatial and temporal smoothing. Indeed, spatial and temporal integration causes the cancellation of the movement along the rays. However, it alone would not give rise to movement within the rings. Thus simple low-level processing cannot explain the illusion.

The new illusory movement patterns support our theory about the construction of the early visual system for the purpose of three-dimensional motion estimation and segmentation, which is as follows.

We assume that when viewing the images and fixating on a point, some small eye movements take place which at the earliest stages cause the generation of a flow field. The motion measurements are perpendicular to the rays, as shown in figure 13*a*. At the next stages of the motion pathway there are neurones having increasingly large receptive fields which are tuned to motion signals that form parts of coaxis and copoint patterns.

As described earlier, various layers of cells participate in the process of evaluating motion patterns, and at the final stage an estimate of three-dimensional motion is obtained. Each of the neurones involved receives flow field information from earlier neurones, when available, and performs a spatial and

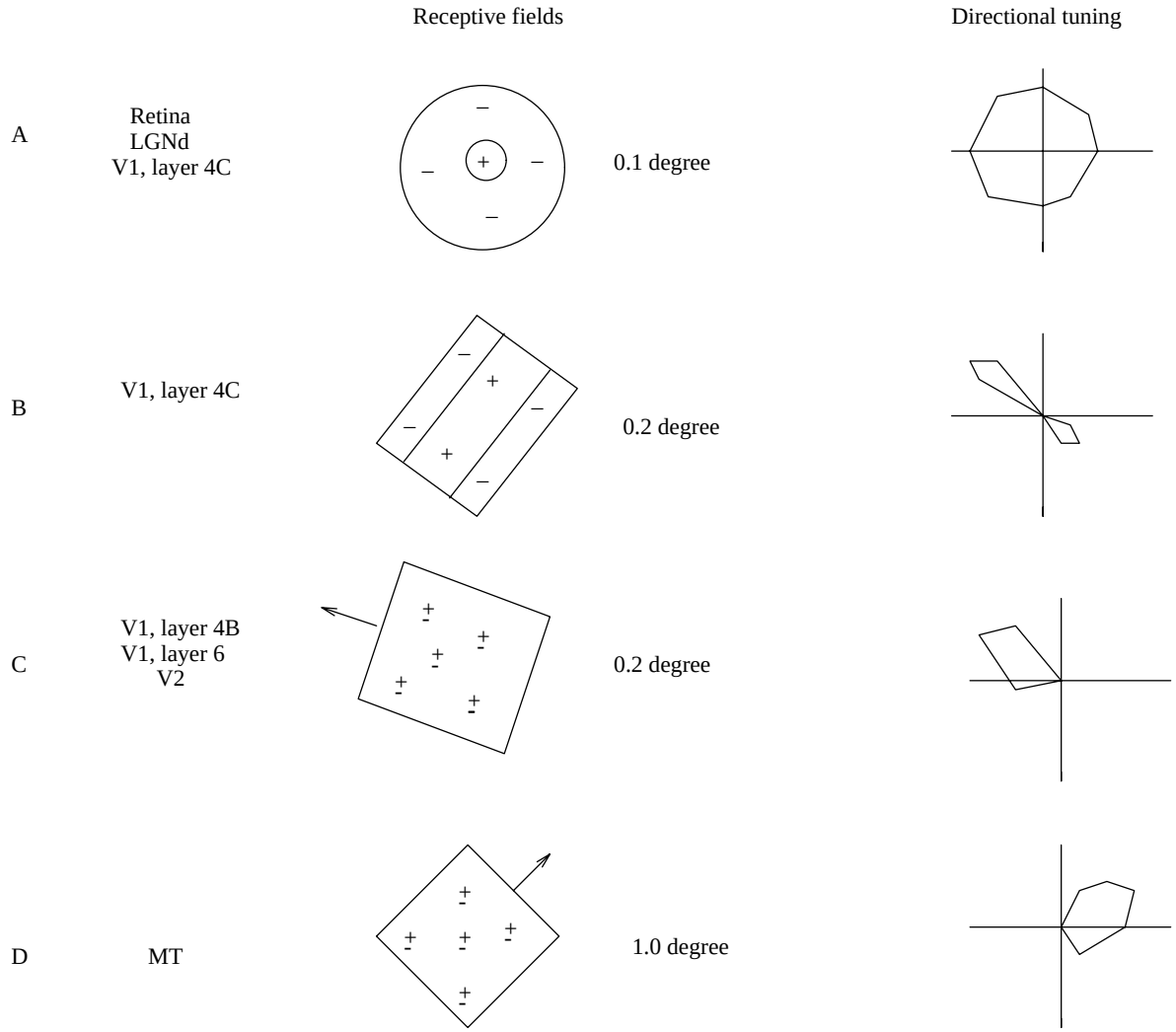


Figure 11. The spatial structure of visual receptive fields and their directional selectivity at different levels of the motion pathway (from Movshon 1990): the spatial scales of the receptive fields (0.1°, etc.) listed here are for neurones at the centre of gaze; in the periphery these dimensions would be larger. The polar diagrams illustrate responses to variation in the direction of a bar target oriented at right angles to its direction of motion. The angular coordinate in the polar diagram indicates the direction of motion and the radial coordinate the magnitude of the response.

temporal integration or smoothing. The system evaluates a large number of coaxis and copoint fields. A set of neurones is devoted to each field, with the single neurones covering certain areas of the image and their receptive fields overlapping. For every illusory figure there should exist exactly one set of neurones responsible for evaluating the related vector field or a very similar one.

When exposed to one of the illusory figures, the receptive fields of some of the cells in the set will cover both an area containing the rays and an area containing the homogeneous part. During the process of pattern evaluation due to the smoothing these neurones create information within the homogeneous parts.

After the neurones with the largest receptive fields involved in pattern matching have been excited and have produced three-dimensional motion estimates (which, however, mean nothing to the stationary observer), the estimates are fed back to the earlier cells, which then have to perform the exact localization of moving objects.

At this stage, temporal integration must cancel out the flow information within the areas covered by the rays. It, however, does not cancel out the motion within the homogeneous regions since the responses from the neurones there do not contradict an existing motion in these areas. Since the system's task, during this processing stage, is to accurately segment the scene, the edges of the homogeneous regions are in some way perceived as motion discontinuities, motion within the homogeneous regions is reinforced, and as a result a strong motion in one direction is perceived. As the small eye movements change sufficiently to change the directions of the elementary motion measurements, the direction of motion in the homogeneous areas also changes.

As mentioned before, our explanation of segmentation is still at a high level. In the explanation given, we concentrated only on motion processes, in particular, three-dimensional motion interpretation and motion segmentation. Recently, neurones believed to be involved in shape processing have been identified which respond to patterns similar in structure to the

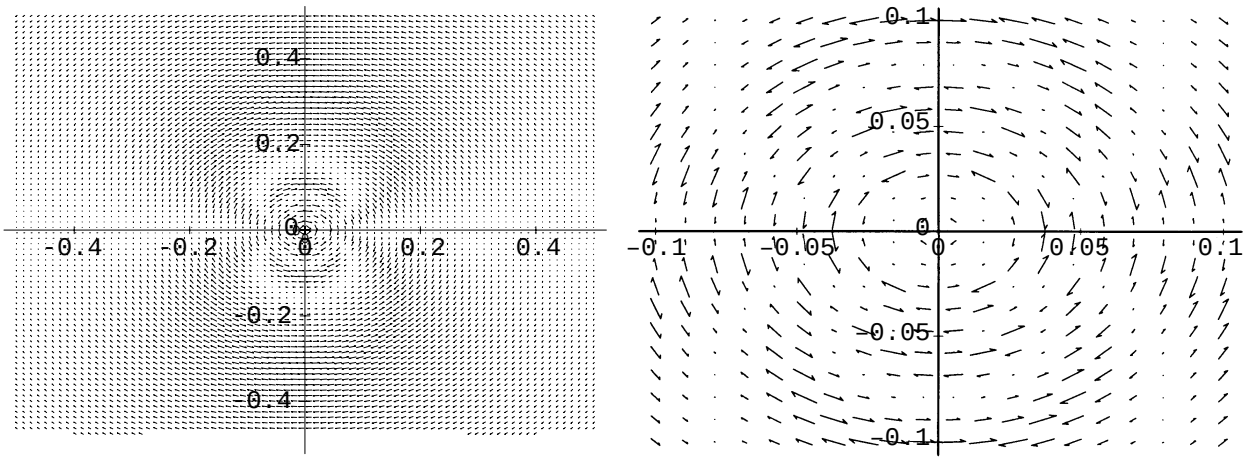


Figure 12. Motion measurements due to small translational eye movements in MacKay rays. On the right is an expanded view of the centre of the flow field; the spatially frequent changes in the velocity constitute the shimmer in the centre of the image.

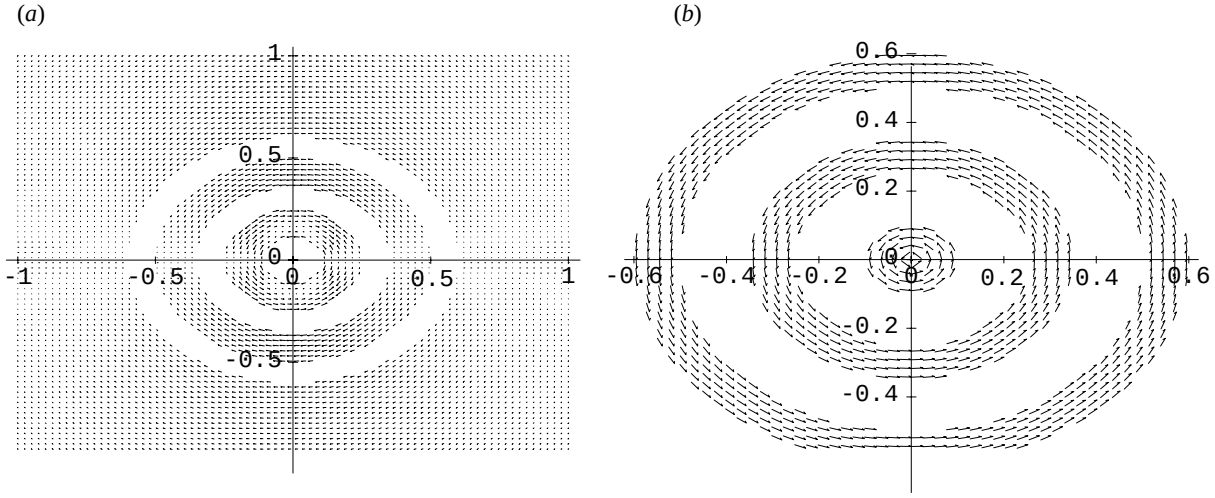
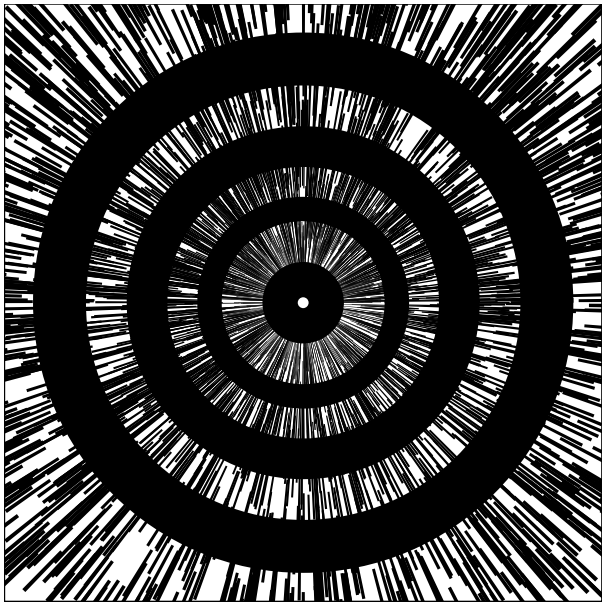


Figure 13. (*a*) Motion measurements due to small translational eye movements in *Enigma*. (*b*) Perceived flow field in Leviant illusion.

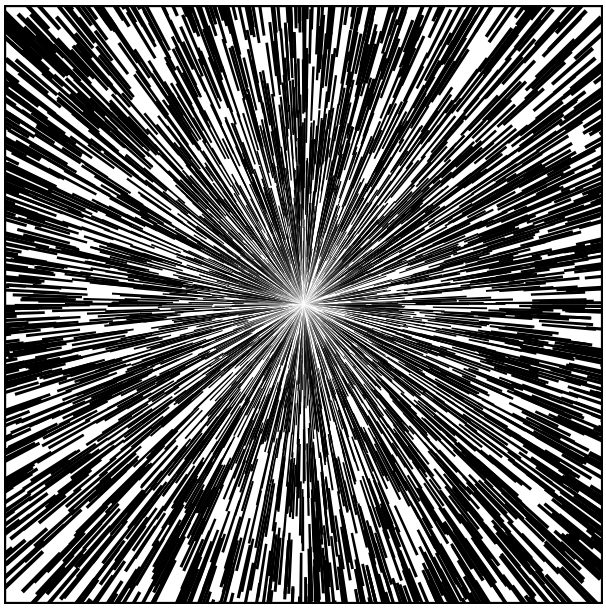
coaxis and copoint fields (Gallant *et al.* 1993). It may be the case that other processes devoted to the analysis of shape, and also to the analysis of colour or intensity, play an additional role in the creation of the illusion. As a matter of fact, intensity influences the strength of the illusion. As already observed by Leviant (1996), the illusory movement appears weakest for black or white bands and strongest for intermediate shades of grey.

The explanation we provided concentrated on the essential computations, and we have performed tests with additional figures to obtain support for the steps of the computational theory as follows. First, in figure 14*a* we replaced the regular radial line pattern with a pattern of random radial line segments. The perception of illusory movement still occurs, although some viewers report it to be of a weaker nature. This figure provides evidence that it is not cells tuned to very regular patterns (i.e. rays at equal distance and rings perpendicular to these rays) which are responsible for the illusion. Instead, it supports the theory that single flow values perpendicular to the rays are computed which at later stages are conglomerated. Together with figure 14*b*, which does not produce the MacKay effect, it also provides evidence that the processes responsible for the *Enigma* illusion must be different from those responsible for the MacKay illusion. Figure 14*b* consists of random line segments only. As can be verified by examination, these lines do not give rise to a MacKay effect; figure 14*a*, on the other hand, does give rise to motion perception within the rings.

(2) Our theory stipulates that within the system there exist neurones having increasingly large receptive fields which piece together the parts of the global motion patterns. If this is the case, weak perceptions of movement should occur within image patches if a figure is viewed which produces motion vectors locally similar to a coaxis or copoint field. This is indeed the case, as can be verified by viewing figure 15. Also, some viewers report very weak, localized movement within the sinusoidal bands of figure 10, particularly in the parts of the bands which are nearly straight.



(*a*)



(*b*)

Figure 14. Replacing the regular radial lines with random line segments does not affect the rotary motion in (*a*) but causes the MacKay effect to disappear in (*b*).

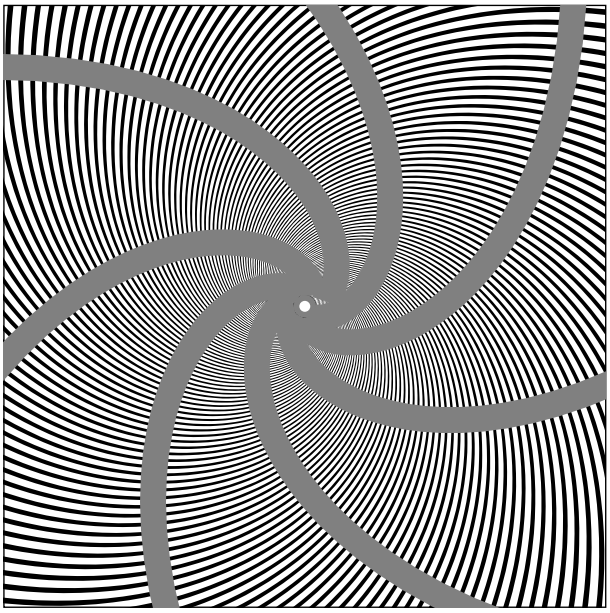


Figure 15. The spiral pattern is locally similar to an illusory pattern due to a coaxis field. A weak perception of movement occurs locally.

5. CONCLUSIONS

A computational explanation of the illusory movement experienced when viewing the *Enigma* figure has been presented using a model for visual motion interpretation developed earlier (Fermüller 1995; Fermüller & Aloimonos 1995*a*, *b*, 1997). This model is consistent with all available neurobiological evidence. The underlying idea is that patterns of retinal motion measurements are formed from the visual flow field; these patterns encode three-dimensional information in a simple manner. The structure of one of these patterns is related to the structure of the *Enigma* figure. The principle of this relationship was used to create, based on other patterns, new figures that also exhibit illusory movement.

Regarding the debate taking place (Gregory 1995; Zeki 1994, 1995) over whether the illusion is due to only high-level processes or only optical causes, our analysis suggests that weak retinal motion signals due to optical causes constitute the triggering of the illusion; however, the major reasons for this illusory movement are high-level processes accessing the retinal motion measurements for the purpose of three-dimensional motion interpretation. We agree with Zeki that examples like this, where the brain appears to go beyond the information given, can give insight into the processes of visual perception.

For example, the computational models proposed in this paper suggest a number of neurobiological experiments aimed at unravelling the functional architecture of the motion pathway with regard to the interpretation of three-dimensional information from visual motion. Our analysis suggests a working model for the functional architecture of the motion pathway which, from a computational standpoint, is optimal. The model explains processes involved in three-dimensional motion estimation. At this time, though, the interplay between three-dimensional motion estimation, segmentation, and also shape and colour processing, is understood only at a high level. We believe that it is up to the synergistic efforts of empirical and computational scientists to unravel the mysteries of these computational processes. In our opinion, a good first step in that direction should be a hunt for neurones in the motion pathway that represent parts of some or all of the coaxis and copoint motion-sensitive orientation fields.

Special thanks to Sara Larson for her editorial and graphics assistance. The support of the Office of Naval Research under Grant N00014-96-1-0587 is gratefully acknowledged.

REFERENCES

Adelson, E. & Bergen, J. 1985 Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am.* A **2**, 284–299.

Aloimonos, Y. (ed.) 1993 Active perception. *Advances in computer vision*. Hillsdale, NJ: Lawrence Erlbaum.

Andersen, R., Snowden, R., Treue, S. & Graziano, M. 1990 Hierarchical processing of motion in the visual cortex of monkey. In *The brain* (*Cold Spring Harbor Symp. on Quantitative Biology*), vol. 55, pp. 741–748. Cold Spring Harbor Laboratory Press.

Duffy, C. & Wurtz, R. 1991*a* Sensitivity of MST neurons to optical flow stimuli. I. A continuum of response selectivity to large field stimuli. *J. Neurophysiol.* **65**, 1329–1345.

Duffy, C. & Wurtz, R. 1991*b* Sensitivity of MST neurons to optical flow stimuli. II. Mechanisms of response selectivity revealed by small field stimuli. *J. Neurophysiol.* **65**, 1346–1359.

Fermüller, C. 1995 Passive navigation as a pattern recognition problem. *Int. Jl Computer Vision* **14**, 147–158.

Fermüller, C. & Aloimonos, Y. 1995*a* Direct perception of three-dimensional motion from patterns of visual motion. *Science* **270**, 1973–1976.

Fermüller, C. & Aloimonos, Y. 1995*b* Qualitative egomotion. *Int. Jl Computer Vision* **15**, 7–29.

Fermüller, C. & Aloimonos, Y. 1997 On the geometry of visual correspondence. *Int. Jl Computer Vision.* (In the press.)

Gabor, D. 1946 Theory of communication. *J. IEE '93*, part III, 429–457.

Gallant, J., Braun, J. & Essen, D. C. V. 1993 Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* **259**, 100–103.

Gregory, R. L. 1993 A comment: MacKay rays shimmer due to accommodation changes. *Proc. R. Soc. Lond.* B **253**, 123.

Gregory, R. L. 1995 Brain-created visual motion: an illusion? *Proc. R. Soc. Lond.* B **260**, 167–168.

Grzywacz, N., Watamaniuk, S. & McKee, S. 1995 Temporal coherence theory for the detection and measurement of visual motion. *Vision Res.* **35**, 3183–3203.

Lagae, L. 1991 A neurophysiological study of optic flow analysis in the monkey brain. Ph.D. thesis, Leuven, Belgium.

Leviant, I. 1996 Does 'brain-power' make *Enigma* spin? *Proc. R. Soc. Lond.* B **263**, 997–1001.

MacKay, D. 1957 Moving visual images produced by regular stationary patterns. *Nature* **180**, 849–850.

MacKay, D. 1958 Moving visual images produced by regular stationary patterns. *Nature* **181**, 362–363.

MacKay, D. 1961 Interactive processes in visual perception. In *Sensory communication* (ed. W. Rosenblith), pp. 339–355. Cambridge, MA: MIT Press.

Mon-Williams, M. & Wann, J. 1996 An illusion that avoids focus. *Proc. R. Soc. Lond.* B **263**, 573–578.

Movshon, A. 1990 Visual processing of moving images. In *Images and understanding* (ed. H. Barlow, C. Blakemore & M. Weston-Smith), pp. 122–137. Cambridge University Press.

Orban, G., Lagae, L., Verri, A., Raiguel, S., Xiao, D., Maes, H. & Torre, V. 1992 First order analysis of optical flow in monkey brain. *Proc. Natn. Acad. Sci. USA* **89**, 2595–2599.

Reichardt, W. 1961 Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In *Sensory communication* (ed. W. Rosenblith), pp. 303–317. Cambridge, MA: MIT Press.

Reichardt, W. 1987 Evaluation of optical motion information by movement detectors. *J. Comp. Physiol.* **161**, 533–547.

Snippe, H. & Koenderink, J. 1994 Extraction of optical velocity by use of multi-input Reichardt detectors. *J. Opt. Soc. Am.* A **11**, 1222–1236.

Tanaka, K. & Saito, H.-A. 1989 Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells illustrated in the dorsal part of the medial superior temporal area of the macaque monkey. *J. Neurophysiol.* **62**, 626–641.

van Santen, J. & Sperling, G. 1984 Temporal covariance model of human motion perception. *J. Opt. Soc. Am.* A **1**, 451–473.

Zeki, S. 1993 *A vision of the brain*. Oxford: Blackwell.

Zeki, S. 1994 The cortical *Enigma*: a reply to Professor Gregory. *Proc. R. Soc. Lond.* B **257**, 243–245.

Zeki, S. 1995 Phenomenal motion seen through artificial intra-ocular lenses. *Proc. R. Soc. Lond.* B **260**, 165–166.

Zeki, S., Watson, J. D. G. & Frackowiak, R. S. J. 1993 Going beyond the information given: the relation of illusory visual motion to brain activity. *Proc. R. Soc. Lond.* B **252**, 215–222.

Received 20 January 1997; accepted 12 February 1997