Ratios of Template Responses as the Basis of Semivision

G. A. Horridge

Phil. Trans. R. Soc. Lond. B 1991 **331**, 189-197
doi: 10.1098/rstb.1991.0007

Ratios of template responses as the basis of semivision

G. A. HORRIDGE

Centre for Visual Sciences, Research School of Biological Sciences, Australian National University, Box 475 P.O. Canberra, ACT 2601, Australia

SUMMARY

The template model starts with a layer of receptors that in the case of vision are leaky detectors or counters of photons. In many animals, the ratio of the responses of a few spectral types is the basis of colour vision irrespective of intensity. Ratios of template responses are now introduced as the basis of form discrimination. In insects, the second-order neurons on the visual pathway appear to detect temporal contrast at the spatial resolution of the retina. At the next level, in the optic medulla, we find a large number of small local neurons in a column on each visual axis. The template theory is a hypothesis about how the above system functions. All possible combinations of positive, indeterminate or negative temporal contrast are considered, at two adjacent visual axes at two successive instants, giving 81 possible local templates. These templates are therefore phasic detectors of all the possible spatiotemporal contrast combinations. Some of the template responses indicate polarity of edge, flicker, or direction of motion and other abstracted features of the stimulus pattern with the maximum spatial and temporal resolution. The ratios of numbers of template responses, in higher fields at a higher level, yield quantitative measures of the qualities of edges independently of the number of edges, but taking ratios causes a corresponding loss of the spatiotemporal resolution and the pattern within each field. Templates respond to transients without computation, are readily modified or selected in evolution and can be simulated in artificial vision.

INTRODUCTION

Semivision is that kind of vision which is inferred from visual behaviour and electrophysiology of lower animals and in which there is no suggestion of a conscious visual world or categorization of objects, but excellent visual discriminations of form and pattern nonetheless. An aim in the study of semivision is to incorporate the principles from natural vision into artificial seeing systems at about the same level of performance.

SEMIVISION

Evolution of visual processing

If visual systems have evolved slowly in different animal groups as a result of trials of many diverse bits of circuitry, this process must have been governed by the detailed nature of the visual images, by the nature of the tasks in visual behaviour, and also by the geometry of the three-dimensional world as it is projected upon the retina of a moving eye, but there is also a historical aspect in that new circuits evolve in the context of what is already there.

Among lower animals such as medusae and tubiculous worms, there are excellent-looking eyes with a lens and a retina of many small photoreceptors, but no sign of visual behaviour that would require more than the simplest visual processing. The most that these eyes do is to detect the least motion of the smallest moving shadow, possibly with the angle on the eye at which such an event occurs, and they signal a command to stop, turn or withdraw. Nevertheless, such a task requires a high spatial sampling frequency which in turn depends upon excellent optics and a large array of small photoreceptors.

The next stage, reached by crustaceans, insects, spiders and some molluscs, allows the animal to sway its head, move through a three-dimensional world or stay still in a scene full of movement, but to detect a small unfamiliar movement against this background. A similar level of complexity of sensory processing occurs in other modalities; for example, an earthworm crawling over a rough surface responds at once to a novel mechanical stimulus. Numerous studies of sensory mechanisms have revealed a rapid synaptic habituation to any regularly repeated stimulus, and in the visual system we can readily imagine a similar adaptation to constant contrast. The other known mechanism which detects a local moving contrast on a background of other moving contrasts is to subtract the average stimulation received by large fields from that received by small fields in the same location. Outstanding differences imply that something unusual is present locally. Frequently this mechanism appears as a neuron with a centre/surround field in the domain of intensity, colour, contrast, or motion, and examples abound wherever visual systems have been explored electrophysiologically. The structure, dimensions and form of the neuron fields are intimately linked to the spatiotemporal resolution of the eye (see, for example, Srinivasan *et al.* (1990)) and also to the dimensions,

Phil. Trans. R. Soc. Lond. B (1991) **331**, 189–197
Printed in Great Britain

motion, intensity, colour and contrast levels of the expected features in the visual image. These in turn are related to the task towards which the vision is directed.

Next we come to directional mechanisms, in which the control of direction and stability in locomotion directly feed back into the visual system of the moving animal. The eye generates the flow field by its own motion, but does not need to analyse the whole of this flow field. Righting reflexes keep the animal on an even keel, and locomotion is usually limited to the horizontal plane. At each point on the retina the motion is therefore only one-dimensional, being horizontal from front to back along the sides of the flow field, diverging from a point at the front, and from front to back in the region looking down on the ground below. General systems for moving sensors (see, for example, Baker & Bolles (1989)) are not necessary. As long as a moving insect keeps its normal posture, there is no evolutionary pressure to evolve a visual system that analyses the moving patterns of contrasts into two-dimensional pictures. Of course, to control flight posture in roll and pitch, some directional detectors of motion in the vertical planes are essential. Indeed we find them in insects (see, for example, Goetz (1968)), but there is no reason to suppose that they contribute to two-dimensional pictures. The optomotor neurons can have large fields and they detect the direction of transient motion but there is no evidence to suggest that they are involved in visual discrimination.

Given an eye with good spatial resolution, evolved to detect motion of a small contrast and its angle on the eye, the stage is set for the further evolution of simple mechanisms for the rapid *discrimination* of different contrasts which are moving in the expected directions in different parts of the eye. Visual discrimination of different features implies that different quantitative features are extracted from an array of inputs independently of the pattern or frequency of passing edges in such a way that contrasts of different spatial structure and range are separable into *different line-labelled pathways*, as indeed is demonstrated in visual behaviour. Colour vision already illustrates the principle that discrimination by an ensemble of neurons is independent of stimulus intensity, and we will return to the colour vision model.

The template model of semivision in one dimension

Recently a model of how contrasts in the visual world are discriminated by a moving eye has been described (Sobey & Horridge 1990) and experimental testing has begun on insects (Horridge & Marcelja 1990*a*, *b*, 1991). The first stages of processing are:

(*a*) The photoreceptors of the retina are an array of sampling stations. At each location on the retina the processing of the expected transients is one-dimensional along an expected line.

(*b*) The outside world consists of contrasting edges of which the angles to the direction of apparent motion are not relevant to vision in one dimension at each point. Many of the relevant edges are vertical, and those for landing are horizontal.

(*c*) Intensities at the receptors are translated separately into analogue signals, which are differentiated with respect to time in the second-order neurons, to yield a temporal contrast at each instant on each visual axis in the columns of neurons of the optic medulla. In the model, this is done by subtraction.

(*d*) These positive or negative temporal contrasts in analogue form are then thresholded to yield one of three states on each visual axis at each instant, by the following rule

$$0.008 > 0 > -0.008 \quad (\uparrow) > (-) > (\downarrow), \qquad (1)$$

i.e. a temporal contrast that exceeds 0.8% is taken as increasing (↑); if less than 0.8% it is taken to be decreasing (↓) and if not noticeable there is said to be 'no change' (–). This step gives us a spatiotemporal map with each location identified with one of these three states (figure 1).

(*e*) There are spatiotemporal interactions between inputs from pairs of second-order neurons on two adjacent visual axes at two successive instants, as a model of the numerous small third-order neurons in the columns of the optic medulla.

(*f*) First, from (*d*), we write out all possible pairs of three states, for one instant, namely:

(––), (↓–), (↓↓), (↓↑), (–↑), (↑↑), (↑–), (↑↓) and (–↓).

For every adjacent pair of inputs along the line of motion, *one* of these combinations must exist.

(*g*) Then, for successive instants we can make all possible 2 × 2 spatiotemporal templates from the nine in (*f*). There are 81 templates (figure 2); one of these combinations must turn up at each instant at every visual axis along the line of motion so all the resolved image is coded as spatiotemporal correlations. The relation between templates and the spatiotemporal map of edges seen by a moving eye is shown in figure 1.

(*h*) In this scheme, templates have the same high spatial resolution as receptors otherwise they miss some of the combinations of receptor inputs. Similarly in the time domain, the time constants of the templates must be matched upstream and downstream.

(*i*) Because each visual axis at each instant appears in four templates, if all 81 templates are present, the whole spatiotemporal map of thresholded changes in the input can be covered by taking only every fourth template response. The full mechanism is therefore resilient to noise and damage.

(*j*) A quarter of the templates, those with 3:1 diagonal symmetry (figure 1), are directionally selective, half of them for one direction and the other half for the opposite direction along the expected line of motion (figure 2). Directions are given looking into the template. If the eye scans from side to side, these sets of directional templates are effective alternately.

(*k*) The state 'no change' (–) is a component of templates of several kinds, e.g., ones representing direction such as (↓–/↓↓), and others representing temporal frequency (––/↓↓) and spatial frequency (↓–/↓–). There is a corresponding set without the 'no change' symbols such as (↓↑/↓↓), (↑↑/↓↓) and (↓↑/↓↑), which, as a group, are adequate for semivision, and which can incorporate a 'zero crossing'.

(*l*) The edges of the stimulus are gradients at the

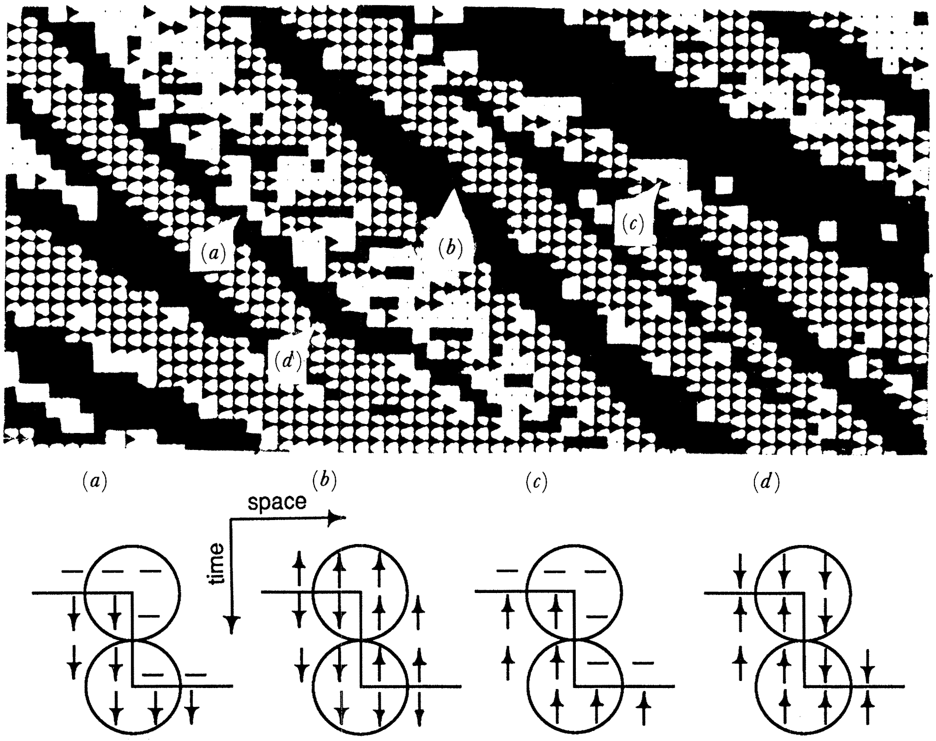

Figure 1. The way that templates correspond to the quantized sampling array, and the digitized state at each point in spatiotemporal coordinates. At the top of the figure is the spatiotemporal map of thresholded contrasts from the central region of figure 3. (■) decreasing contrasts (▶) increasing contrast. (●) no change. (*a*) to (*d*). Four regions of this map further magnified with symbols as in figure 2. Each group of 4 symbols becomes a template. These templates shown have 3:1 structure and diagonal symmetry which fit the corners of the steps in the spatiotemporal representation of an edge moving to the right.

t_1 \ t_0	– –	↓ –	↓ ↓	↓ ↑	– ↑	↑ ↑	↑ –	↑ ↓	– ↓
– –	Colour	B ←	V		W →	V	W ←		B →
↓ –	B →	V Colour Form	B ←						
↓ ↓	V	B →		B →		V		B ←	B ←
↓ ↑			W ←	V		B →			
– ↑	W ←				V Colour Form	W →			
↑ ↑	V		V	W ←	W ←	V	W →	W →	
↑ –	W →					W ←	V Colour Form		
↑ ↓			W →			B ←		V	
– ↓	B ←		B →						V Colour Form

Figure 2. The 9×9 table of 2×2 contrast templates. (↑), increase in temporal contrast; (↓) decrease in temporal contrast; (–) no change; (→) and (←), inferred directions of motion; B, black following white; W, white following black; V, templates useful for measurement of velocity. Templates useful for form or colour are also indicated.

retinal receptors, but are made sharp by yes or no decisions of templates in the central projection. Edges can be detected without a 'zero-crossing' computation. Differences in quality of edge are evident in the groupings of template responses.

(*m*) Only the useful templates are selected in the evolution of the system and therefore much of the spatiotemporal map may generate 'no response', which is not the same as 'no change'.

(*n*) The templates are line-labelled, like neurons, according to what stimulates them. Their responses can be counted and collected locally or globally into specific combinations, so that more complex properties can be detected. All the principles of ensemble coding apply (Maes & Erickson 1984; Maes & Ruifrok 1986).

(*o*) The templates begin the segregation to parallel neuronal systems for direction of motion, polarity of edge, and non-directional motion. Colour, background intensity and effects of the plane of polarization must be carried in other channels.

(*p*) Individual template responses convey little because spatiotemporal contrast templates of only 2 × 2 units are relatively non-specific. Combinations of templates are necessary, otherwise processing would not be distributed.

(*q*) Some templates are rarely called into action and never occur in groups; others commonly respond, with obvious consequences for the further evolution of mechanisms.

(*r*) Templates can be combined together by simple logic to make wider and longer templates for any specific task in predetermined eye regions. The numbers of combinations is limited by defining the visual tasks.

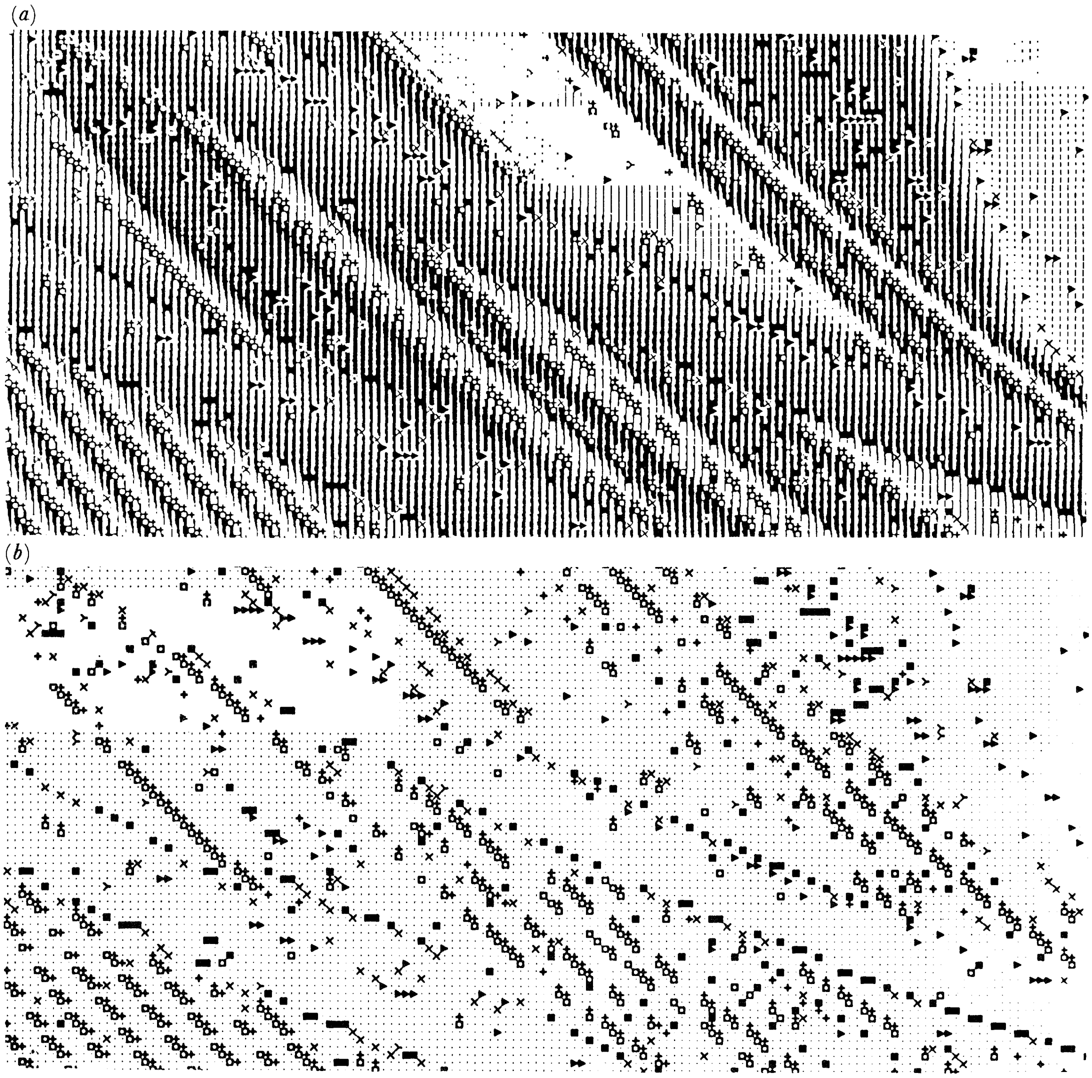

Figure 3. (*a*) Successive scans (plotted downwards in time) of a natural scene by a single line of 150 pixels (plotted horizontally) to give a spatiotemporal map of intensity of the moving one-dimensional scene. The responses of the following templates from this map are then superimposed. ■, (↓↓/↑↑); +, (↑↑/↓↑); ≻, (−↑/−↑); □, (↓↑/↓↓); ▶, (↑↑/−−); ×, (↑−/↑↑). (*b*) The same map with template responses only.

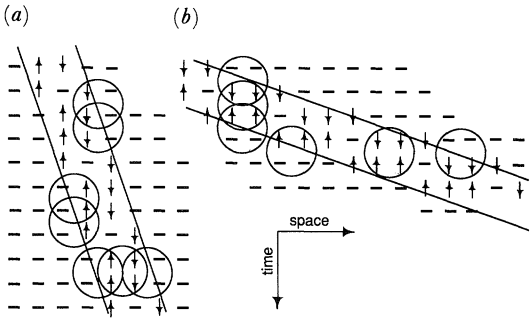

Figure 4. Motion of a narrow dark bar from left to right (*a*) slowly (*b*) faster, showing different ratios of (−−/↓↓) or (↑↑/−−) to (↓−/↓−) or (−↑/−↑) and the location of (↑↓/↑↓) and (↓↓/↑↑) which indicate a polarity change in the appropriate context.

Ratios of counts of template responses

The first processing layer, above, feeds into units in the model with medium size fields that allow discriminations of more complex features no matter how frequently they are scanned, as in colour vision. For each of these higher level units, the arrangement of features within its field is lost anyway. The way to process the template responses, therefore, is to count them like photons in photoreceptors of different spectral peak, and then make discriminations on the basis of relative numbers of different template responses within each field. We accept that in the formation of higher fields the form and resolution of pattern is lost but discrimination of other kinds is retained. For example, differences in the angular velocity of edges or bars across the eye, as a measure of range, are indicated by the relative numbers of some non-directional templates such as (−−/↓↓) and (−↓/−↓) representing the ratio of temporal gradient to spatial gradient (figure 4) so that velocity is just another edge quality in a spatiotemporal field.

To give maximum discrimination by higher level mechanisms, the field sizes over which ratios are counted must be about the same size and duration as the feature qualities they are to detect otherwise features are missed, and not larger than the distance

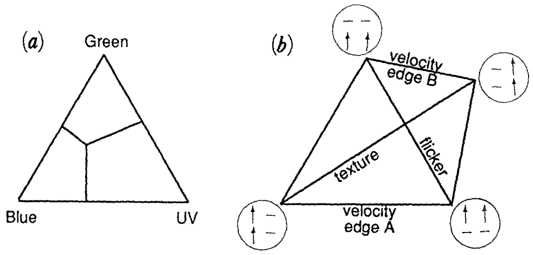

Figure 5. (*a*) Colour triangle; any point in the triangle represents the ratios of three primary colours. (*b*) For groups of templates there is a similar representation which can be expanded to any number of dimensions. Non-directional templates (↑–/↑–) and (–↑/–↑) indicate light following dark either way; (↑↑/––) and (––/↑↑) indicate the end and beginning of a temporal change.

between features, otherwise more than one feature is included in a field. With the taking of ratios, the 2 × 2 templates give a reasonable rejection of noise or irrelevant pixel combinations but avoid high resolution processing by more specific templates, which are much more difficult to evolve to match the relevant features of the visual scene. Having assumed that the scene consists of contrasting edges that move predictably across the retina, and having used the three symbols (↑) (↓) and (–), this data-oriented algorithm delivers the angular direction on the eye of increasing and decreasing contrasts, no change, moving edges of various qualities and regions of differing textures. The model is colour blind and does not preserve the spatial arrangement of the features detected by templates at the high resolution of the retina. Nor can this mechanism by itself provide specific templates that recognize flowers and other insects at a distance. However by subsequently bringing together the 2 × 2 templates in refined ratios followed by logical 'and' combinations, we could make more complex filters that respond instantly and as specifically as we please to particular visual targets of great selective value, irrespective of pattern. Counting template responses over local spatiotemporal regions can be represented in the same way as colour triangles or tetrahedra (figure 5), or in more dimensions, or the excitation can continue in parallel neurons. Such a scheme, copying the ensembles of neurons at every stage, appears to be a realistic model because electrophysiology reveals only numerous higher neurons which still function in groups right through to the motorneurons.

In contrast to templates, neurons overlap, adapt and are more variable. Also, they have more states in that their responses may change more smoothly in time. Templates may be neurons, several neurons or part of a neuron, but they exemplify parallel processing which looks rather like the processing mechanism in a real nervous system. Unlike neurons, templates responses are yes/no decisions based on three possible states at each pixel, and have 2 × 2 structure which originates in the quantized sampling of a continuous moving image. Having more states or continuous sampling does not alter the basic principles. The consequences of adaptation in templates remain to be investigated.

Semivision of objects in two dimensions

We can now ask what is the natural and most economic way to combine templates for operating in two spatial dimensions. How do templates evolve in that direction.

The premise for semivision in one-dimensional motion was that the visual image is composed of edges moving along an expected line. The equivalent premise for two dimensions is that the visual image is composed of corners which also move along an expected line. The simplest two-dimensional spatiotemporal templates have a 2 × 2 × 2 structure (figure 6) but we immediately see two difficulties. First, there are 6561 of these templates, so we may be limited to a few that are the most effective for any given visual task in a typical scene. Only natural evolution or implementation of trials can make that choice. Therefore we implemented these 2 × 2 × 2 templates on natural scenes, using a frame grabber to hold many successive camera images. The result was clear but disappointing. There are so many 2 × 2 × 2 templates that the responses of each one are scattered infrequently across a moving scene in apparent confusion. Templates with 3 × 3 visual axes in space make even less sense. That approach appears to be blocked by the combinatorial explosion, there being 387,420,489 templates of the 2 × 3 × 3 type. Another obvious problem is that these two-dimensional templates give unwanted responses to one-dimensional patterns that happen to fit. In brief, if we make the templates sufficiently specific for two-dimensional features, we have too many of them. Exploring the practicality of detecting two spatial dimensions simultaneously clearly shows the limitations of on-line vision with a very small brain or light-weight mobile computer. The data-oriented algorithm must use very simple templates (table 1) which function in groups.

Having explored two-dimensional templates, we look again at the situation in insects, where the directions of the flow lines on the eye are predictable except in the parts looking forward. Therefore we can

Table 1. *The effect of the combinatorial explosion assuming three states. With one-dimensional 2 × 2 templates there is a reasonable chance (10%) of a possible combination of inputs being a useful template*

(As the template increases in size and dimensions, the number of possible combinations increases so rapidly that the appropriate template would never be found, either in evolution or normal function. An alternative is to follow the example of colour vision, and count ratios of responses in few templates with a corresponding loss of spatial arrangement and temporal resolution.)

no. of spatial dimensions	1	1	1	2	2	2
template size	2 × 2	2 × 3	3 × 3	2 × 2 × 2	2 × 2 × 3	2 × 3 × 3
no. of possible combinations of each visual axis	81	729	19683	6561	531441	3.87×10^8
minimum number for directional sense	4	4	4	32	32	32
possible number employed (double the minimum)	8	8	8	64	64	64
useful fraction	10^{-1}	10^{-2}	4×10^{-4}	10^{-2}	10^{-3}	5×10^{-6}

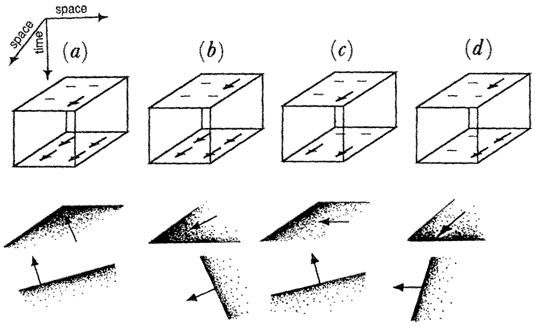

Figure 6. (*a*), (*b*) Two $2 \times 2 \times 2$ templates that detect motion of a corner in a direction diagonal to the corner, and (*c*), (*d*) two that detect motion of a corner along one of its own edges. Templates with two spatial dimensions are too numerous to be useful, not sufficiently specific to detect features in real scenes, and respond to some one-dimensional stimuli, as shown in the lower line.

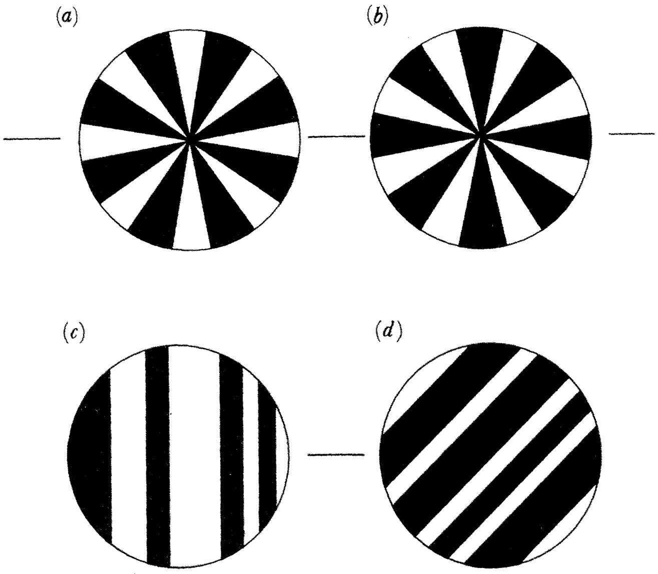

Figure 7. Data relating to the question of two-dimensional templates in bee vision. Bees can distinguish between (*a*) and (*b*) if allowed to approach closely to the reward hole at the centre. One pattern is displaced 22.5° relative to the other. From a distance, bees can distinguish the angular difference between (*c*) and (*d*) irrespective of which of many random patterns is presented, but they cannot distinguish these two patterns when shown at the same angle. ((*a*, *b*) after Wehner 1981; (*c*, *d*) after van Hateren *et al.* 1990 and unpublished results of Dr Srinivasan.)

postulate that vision is one-dimensional but within the same area of the eye there can be horizontal and vertical templates of the 2×2 form illustrated in figure 1. We already know that insects have directional unit motion detectors for vertical as well as horizontal planes which feed separately into large-field neurons (Hausen & Egelhaaf 1989) but there is no suggestion that horizontal and vertical units combine on every visual axis to create a simultaneously two-dimensional system. As an example of how templates are probably combined let us propose that within a medium-size field looking forwards the relative numbers of responding templates in vertical and horizontal planes can be counted during a horizontal scan by a bee. The ratio gives the average angle of inclination of edges within the field but pattern and temporal sequence is sacrificed in favour of measurement of angle. This is no more difficult in principle than discrimination of a colour by ratios. Now let us consider the actual performance of bees making the discriminations in figure 7, on a vertical surface. They can distinguish the small difference in position of the patterns in 7*a*, and 7*b* if they are allowed a close inspection, perhaps by using some trick. From a distance they can discriminate a difference of 45° in the angle of the random stripes in figure 7*c* and *d*, but they cannot tell the difference between the two patterns in figure 7*c* and *d* if they have the same angle (van Hateren *et al.* 1990; M. V. Srinivasan, unpublished data).

This example shows how discrimination of one feature (angle of tilt) can occur with loss of resolution of another (form or number of edges). Motion of spots and corners, as well as many other essentially two-dimensional features, can also be detected by ratios of one-dimensional template responses. In fact, to overcome the problem that all template responses are 'event-driven', i.e. they respond to the temporal frequency of their feature, we must take a *ratio* to obtain discriminations independent of pattern or frequency of presentation. Then, we can take a logical 'and' of just those template combinations required, whether or not they are in two *or more* dimensions. Neurons, of course, excel at facilitation and inhibition, which are ways of implementing the 'and' integration. So far as I am aware, there is no experimental evidence to contradict the view that insect visual processing has only one-dimensional templates which are integrated in the same way as in colour discrimination. We now need a period of experiment in which to test these ideas by behavioural discriminations and electrophysiological search for the components at neuron level.

The stages of evolution of templates

Parallel distributed processing by templates is a mechanism that starts evolving easily and continues by augmentation. In a group of numerous small neurons behind an evolving eye, let us suppose that connections form, with short-lived rapid-acting transmitter at some synapses, and more slowly-acting persistent transmitter at others to provide the short-term delay between successive instants. At first, movement detectors have the most survival value. Later, any one of a large number of connections is likely to fit with some commonly occurring feature in the moving image. It doesn't have to be one of the 2×2 templates: almost any new neuronal connection will respond to something. Slowly, therefore, and piece by piece, templates can be evolved. If there already exists a circuit that effectively counts response ratios within higher-order fields or makes a logical 'and', then neurons are available to take over these functions as new templates appear, with corresponding loss in spatiotemporal resolution. This evolution of circuitry is a process that is never complete, as new functions appear or become redundant. Eye tremor, or scanning, coupled with counts of template responses over longer times, to some extent takes care of undersampling by individual templates.

The size of the higher-level fields within which counts are made must also be a compromise that depends on choice of template and on the visual task. Fields that are too small or too brief will catch too few template responses whereas fields that are too large sacrifice too much of the spatial and temporal resolution of the retina.

How insects avoid classical problems of vision

Some of the classical problems of vision are solved by insects, possibly by the mechanisms outlined in the template model.

(i) Insects do not categorize shapes in pictures and there is no evidence that they even recognize individual objects as being separate from each other or as having outlines, except by parallax and one-dimensional relative motion caused by predictable self-motion of the eye.

(ii) There is no evidence that insects *locate the position* of edges with the resolution of the retina. Therefore all operations, such as zero crossings, to locate edges accurately are irrelevant.

(iii) Insect vision has evolved to match the flow field caused by normal locomotion. Apart from optomotor reflexes and looking straight ahead, this flow field presents one-dimensional motion at each location on the retina, and so economises on templates. With predictable motion, ratios or conjunctions of template responses are adequate for measuring the qualities of moving contrasts. General systems (Baker & Bolles 1989) are not necessary.

(iv) The 'aperture problem' in vision arises because the motion of a straight edge as seen in a restricted field can only be seen as motion at right angles to the edge. The motions of the different edges around an object's outline must therefore be put together to calculate the average vector of motion of all the separate fields. This problem does not arise with a one-dimension motion of a flow field which is predictable at each point on the eye, and is not relevant when spatial arrangement is sacrificed.

(v) The 'smoothness constraints' in vision assume that bounding edges are continuous, that velocity is constant over small areas of the image, and that objects have rigid boundaries. These assumptions simplify the computation of two-dimensional motions but are not relevant to one-dimensional motion in an expected flowfield.

(vi) In contrast to (4) and (5), the maximum resolution of discontinuities in the flow field is significant in insect vision because the motion of contrasts across the background gives information about the separateness of objects and their range. Parallax is a reliable indicator in one-dimensional vision in a predictable flow field. Coupled with lateral scanning, the mere identification of parallax at each angle on the retina perhaps gives sufficient semivision for manoeuvring in flight, without recovering the whole three-dimensional structure.

(vii) In view of (6), the insect visual processing does not need the premise that the world is rigid, and, in fact insects which interact together visually while in flight cannot make that assumption. Perhaps all they see is the range in each direction, relative motion of edges over background and some qualities of edges and textures as measured by ratios.

(viii) The enormous numbers of combinations of inputs, particularly if there can be many states of each, sets an impossible task for a complete system of analysis of unexpected patterns in any orientation. Insects avoid this 'combinatorial explosion' by rapid divergence of the image into sets of collaborating templates that are as simple and appropriate as possible. After this layer there is convergence upon groups of higher-order neurons which are strongly task-oriented so that only useful combinations of a few relevant template responses are counted.

(ix) Electrophysiology suggests that, as in colour vision, the diversity of the first layer of templates is not great, and that to compensate for this, the subsequent neuronal pathways act in groups at all levels.

(x) Insect visual behaviour gives the impression that there is a very limited flexibility in appropriate combinations of more specific template responses at a higher level. Insects seem to be mainly hard-wired for particular tasks such as detection of a mate, prey or flower by groups of neurons, exactly as in their chemosensory behaviour. Diversity of behaviour is achieved by having a diversity of insects.

(xi) Despite (ix) and (x), learning exists in visual behaviour but again it is strictly task-oriented within a limited repertoire, such as landmarks, and colour for bees. Learning a two-dimensional shape on a vertical surface by bees is possible but limited, and easily confused by disruption of the pattern.

EXPERIMENTAL ANALYSIS

The original premise, that the mechanisms of vision which evolved are the ones useful for the animal's own motion through natural scenes, together with the idea that templates respond each to their own specific combination of contrasts, implies that the experimental analysis of an unknown natural visual system must arrive somehow at the use of the correct stimulus patterns. Incorrect test stimuli yield misleading results. Some examples may illustrate this.

Flashing of a point source light while recording from optic lobe neurons yields 'on', 'off' and 'on-off' units, but this conclusion depends on the stimulus. This result can be extended to the careful measurement of time constants and different kinds of spatiotemporal interaction between 'on' and 'off' but the results do not show the optimum stimuli, and over the past 40 years have not generated much understanding of how the neurons collaborate together. White noise analysis is better for recording from a simple template but there is inadequate stimulus power to drive all the combinations needed to identify complex templates. Similarly, stimulation by moving sine-wave patterns is inappropriate where the mechanisms are mainly phasic and include thresholding. By use of a stimulus such as a light spot or a striped pattern which always contains equal numbers of black/white and white/black edges, the results inevitably have the large second harmonic, compatible with a theory such as multiplication or rectification that is independent of edge polarity, as done by Emerson *et al.* (1987). Another set of interpretations threatened by the new theory relate to tests with the 'phi-phenomenon' in which a point source or a small spot jumps a short distance to mimic motion. In contrast to the interpretation given by Chubb & Sperling (1989) and many previous workers, the separate templates with a threshold look at the left and right edges of the slightly blurred patches on the retina and the interpretation by the visual system as a whole depends on whether templates for edges of opposite polarity are present and on the geometry of the edges of opposite polarity. One cannot assume that a single-channel computation is done by the nervous system. The interpretation of neuron responses depends on the premise that in natural vision the single receptors are never stimulated alone by flashing stimuli. The only way to tackle the analysis of natural visual systems is to record from identifiable neurons at every level, especially at the major divergences into parallel lines, and work towards an optimum stimulus for each neuron.

Another premise, that neuron responses act in groups, and that any distributed parallel mechanism must be composed of units that are inadequate alone, means that neurons must be individually identified and also recorded in groups, together with their outputs downline. Unfortunately, it is much easier to construct minimum models than to discover natural visual processing mechanisms.

CONCLUSION

The model presented is not so much a model of the neuron fields, although templates can be compared or equated with inputs that are effective for neurons. Rather, principles learnt from ensemble coding and natural vision have been put together to design the outlines of an artificial seeing system which functions with minimum of computation. First, the template structure was that required to respond to the jump of an edge by one receptor spacing to utilize the full spatial resolution of the eye. The different possible combinations of inputs automatically generated all possible templates. The responses are essentially rapid yes or no decisions to the passing combinations of local contrast in space and time, a shallow computation in many dimensions. In response to the predictable movement of the eye in natural scenes, different templates respond in particular groups. The idea of ratios of responses was drawn from the colour triangle, an idea which can be generalized to any number of dimensions. These mechanisms can be approached by electrophysiology but the theory is essential before we can devise experiments to test whether the qualities of edges are represented by ratios of responses in neuron groups. The template idea fits the anatomy in that there is extensive divergence to many neurons just at the level where the maximum spatial resolution ends. It also agrees with the dependence of nervous processing on having many line-labelled neurons in parallel, and the electrophysiological finding that individual neurons respond to unexpected features such as non-directional motion which do not make sense in isolation. In the model, the image of the outside world is formed in the central projection in many separately line-labelled dimensions, each of which represents a different spatio-temporal combination of contrasts in the image.

The most far-reaching implication however, to my mind, is that the templates generated in parallel are individually inadequate to convey specific information without ambiguity, and therefore are only able to work in groups. They teach us that the fields of single neurons cannot be expected to make sense in isolation. If neurons were highly specific, like complex templates, they would not need to operate in parallel and moreover there would be a very large number of specific options, which means more genetic control and much more time required to evolve them. Templates can be a convenient size only if they function in groups. The same can be said about other levels of complexity in living systems. The division of labour in society saves each of us from having to respond in specific ways to all possible situations but we cooperate within society as a whole. Within the cell, the numerous proteins could not function in isolation. The idea of *emergent* properties has for a century been applied to systems where the properties of the whole are qualitatively different from those of the parts. The template model of vision illustrates exactly how the emergent properties arise because the components of a parallel processing mechanism are necessarily inadequate to carry out individually any of their functions and are dependent upon each other because they have evolved in the context of working together. This generalization implies that self-contained algorithms are not appropriate for a nervous system or for artificial visual processing.

I thank many who have worked on these and related questions in Canberra over the past few years while these ideas have been fermenting. The comprehensive theory has depended on research on the visual behaviour of the bee, much of it done by Miriam Lehrer, on the electrophysiology of insect optic lobes by Ljerka Marcelja over many years, and on computer modeling done recently by Peter Sobey. There are many others besides, working on insect visual processing. Especially, however, I would like to thank my colleague Mandyam Srinivasan for his interest and unfailing support, and for the ideas and critical comments that emerge from our discussions. Numerous drafts of the MS have been typed by Elizabeth Watson. The work on object vision in insects is supported by a grant from the Fujitsu Company of Japan, by funds from the Centre for Visual Sciences and the Centre for Information Science in ANU.

REFERENCES

Baker, H. H. & Bolles, R. C. 1989 Generalizing epipolar-plane image analysis on the spatiotemporal surface. *Int. J. Comp. Vis.* **3**, 33–49.

Chubb, C. & Sperling, G. 1989 Two motion perception mechanisms revealed through distance-driven reversal of apparent motion. *Proc. natn. Acad. Sci. USA*, **86**, 2985–2989.

Emerson, R. C., Citron, M. C., Vaughn, W. J. & Klein, S. A. 1987 Nonlinear directionally selective subunits in complex cells of cat striate cortex. *J. Neurophysiol.* **58**, 33–65.

Goetz, K. G. 1968 Flight control in *Drosophila* by visual perception of motion. *Kybernetik* **4**, 199–208.

Hausen, K. & Egalhaaf, M. 1989 Neural mechanisms of visual course control in insects. In *Facets of vision* (ed. D. G. Stavenga & R. C. Hardie), pp. 391–424. Heidelberg: Springer Verlag.

Horridge, G. A. & Marcelja, L. 1990*a* Responses of the H1 neuron of the fly to jumped edges. *Phil. Trans. R. Soc. Lond.* B **329**, 65–73.

Horridge, G. A. & Marcelja, L. 1990*b* Responses of the H1 neuron of the fly to contrast and moving bars. *Phil. Trans. R. Soc. Lond.* B **329**, 75–80.

Horridge, G. A. & Marcelja, L. 1991 A test for multiplication in insect directional motion detectors. *Phil. Trans. R. Soc. Lond.* B. **331**, 199–204. (Following paper.)

Maes, F. W. & Erickson, R. P. 1984 Gustatory intensity discrimination in rat NTS: a tool for the evaluation of neural coding theories. *J. comp. Physiol.* A **155**, 271–282.

Maes, F. W. & Ruifrok, A. C. C. 1986 Neural coding of salt taste quality in the blowfly. *Calliphora vicina*. II. Ensemble coding. *J. comp. Physiol.* **159**, 89–96.

Sobey, P. & Horridge, G. A. 1990 Implementation of the template model of vision. *Proc. R. Soc. Lond.* B **240**, 211–229.

Srinivasan, M. V., Pinter, R. B. & Osorio, D. 1990 Matched filtering in the visual system of the fly: large monopolar cells in the lamina are optimized to detect moving edges and blobs. *Proc. R. Soc. Lond.* B **240**, 279–293.

van Hateren, J. H., Srinivasan, M. V. & Wait, P. B. 1990 Pattern recognition in bees: orientation discrimination. *J. comp. Physiol.* **167**, 649–654.

Wehner, R. 1981 Spatial vision in arthropods. In *Handbook of sensory physiology* (vol. VII/6C) *Vision in invertebrates* (ed. H. Autrum), pp. 287–616. Heidelberg: Springer Verlag.

Received 15 August 1990; accepted 3 December 1990