Spatial Frequency Tuned Channels: Implications for Structure and Function from Psychophysical and Computational Studies of Stereopsis [and Discussion]

J. P. Frisby, J. E. W. Mayhew, P. E. King-Smith, D. Marr and K. H. Ruddock

Phil. Trans. R. Soc. Lond. B 1980 **290**, 95-116
doi: 10.1098/rstb.1980.0085

Phil. Trans. R. Soc. Lond. B **290**, 95–116 (1980)
Printed in Great Britain

Spatial frequency tuned channels: implications for structure and function from psychophysical and computational studies of stereopsis

By J. P. Frisby and J. E. W. Mayhew
Department of Psychology, University of Sheffield, Sheffield S10 2TN, U.K.

Various psychophysical experiments investigating the role of spatial frequency tuned channels in stereopsis are reviewed and a computational model of stereopsis deriving from these studies is described. The distinctive features of the model are: (1) it identifies edge locations in each monocular field by searching for zero crossings in non-orientated centre–surround convolution profiles; (2) it selects among all possible binocular point-for-point combinations of edge locations only those which satisfy a (quasi-) collinear figural grouping rule; (3) it presents a concept of the orientated and spatial frequency tuned channel as a nonlinear grouping operator. The success of the model is demonstrated both on a stereo pair of a natural scene and on a random-dot stereogram.

1. Introduction

One of the most impressive achievements of the last 15 years or so of vision research has been the psychophysical and neurophysiological investigation of the spatial frequency (s.f.) tuned 'channel' (for reviews, see the paper in this symposium by Campbell, one of the originators of this work). The evidence is now overwhelming that s.f. channels exist in the visual system of many species, including man, and we know a good deal about such things as channel tuning (see, for example, Mostafavi & Sakrison 1976; Wilson & Giese 1977; Wilson & Bergen 1979). Curiously, however, relatively little attention has been given to the question of what functional roles the s.f. channels might serve (Marr's (1976) paper is seminal exception), and even less to empirical tests of such ideas as have been formulated. The channels exist, but what do they do? We have considered how s.f. channels might contribute to texture discrimination (Mayhew & Frisby 1978*c*), but here we review our work on the possible roles served by s.f. channels in stereopsis.

The problem solved by the visual system in achieving stereopsis is easily summarized: it is to compute descriptions from the left and right retinal images, match them correctly binocularly, measure the associated disparities, and thereby build up a depth map of the visual scene. In principle, relatively high-level descriptions such as object (and/or surface) boundaries could be computed and matched. In this case, the visual system might achieve the goal of correct matching by taking advantage of the constraint that an object usually presents roughly similar boundary shapes in the two retinal images. Of course, use of object boundaries could not provide an entirely satisfactory basis for stereopsis because we easily see the depth variations that frequently occur within such boundaries. Moreover, the random-dot stereogram (Julesz 1960) demonstrates conclusively that object contours are not necessary for the stereopsis computation because no objects can be seen in either random-dot stereo half before fusion. Whether monocularly discriminable object contours are nevertheless a *sufficient* basis for at least a limited form of stereopsis remains a controversial question (Ramachandran *et al.* 1973; Mayhew & Frisby 1976; Frisby & Mayhew 1977*a*; Mayhew *et al.* 1977; Frisby & Mayhew 1978), although there is no doubt that monocular cues can play a valuable role in guiding the eye movements required for fusion in certain circumstances (e.g. when disparities larger than those covered by Panum's fusional area are presented; Saye & Frisby (1976); Kidd *et al.* (1979); see also Julesz & Oswald (1978)).

Given the finding that high-level descriptions are not necessary for stereopsis, it is natural to propose that disparity measurements begin with the matching of low-level left/right 'point' descriptions (Julesz 1960). This suggestion, however, immediately provokes the follow-up question: what exactly is a 'point' for stereo combination? Grey level points are intrinsically unsatisfactory because a grey level description does not reliably define a point on a physical surface (Marr & Poggio 1976). Hence, what are required are point descriptions that make explicit changes in reflectance. Of course, given the wide range of textures that can serve as carriers of stereopsis (Julesz 1971), there is almost certainly no single definition of a 'point' that could cope with all of the many different types of reflectance changes that can serve as stereo inputs. In this paper, we review some psychophysical studies that have helped to define what point descriptions are in fact used by the visual system for computing stereopsis, and we describe a computational model of stereopsis whose design has been guided by this psychophysical work.

2. Spatial frequency tuned point descriptions for stereopsis

Julesz & Miller (1975) investigated spatial frequency (s.f.) selective processes in stereopsis by exploring the effects of adding masking noise of carefully controlled spectral composition to bandpass filtered random-dot stereograms. They found that if the spectral content of the noise lay more than two octaves distant from the spectral content of the stereoscopic image, then stereopsis was unaffected. That is, a stereoscopic percept 'drawn' in one s.f. could be seen quite clearly despite the presence of noise from another s.f. band. If, however, the s.f. of the noise overlapped that of the stereoscopic image, then stereopsis was destroyed. They concluded from this result that 'the global stereopsis mechanism (which processes the binocular fusion of random-dot stereograms) utilizes frequency-tuned analyzers, thus these analysers must reside prior to the stage (or at the stage) of global stereopsis'.

Julesz & Miller's ideas were developed in greater detail by Frisby & Mayhew (1977*a*) who were concerned to apply them to the theoretical problems posed by 'rivalrous texture stereograms' (see also Mayhew & Frisby 1976). Figure 1 illustrates this development and makes clear exactly what Frisby & Mayhew meant by an 'independent s.f.-tuned stereopsis channel'. Each such channel receives its inputs from monocular s.f.-tuned analysers which filter the retinal images for particular bands of spectral content. High and low s.f. channels only are depicted in figure 1, but the dotted lines between boxes represent channels tuned to intermediate frequencies. Each monocular s.f. channel can be thought of as a 2D array of analysers, each of which possesses a receptive field whose shape (see profiles in the centre of the figure) determines the s.f. selectivity of the channel. The row of dots within the monocular boxes of figure 1 represent a 1D slice of each 2D array. The 'activity profile' of the analysers in the slice is shown when they are responding to a pattern with a luminance profile as depicted in the retinal image boxes. These activity profiles were obtained from a computer simulation which convolved the input retinal waveform with the high and low s.f. receptive field profiles shown. In neurophysiological terms, excursions of the activity profile above the row of dots is to be interpreted as above-threshold activity in on-centre units, excursions below as activity in off-centre units. Thus each dot can be thought of as representing a pair of cells, one selective for 'brightness' and one for 'darkness' (Jung 1973). The left and right activity profiles for a given s.f. are combined separately for the purposes of both local and global stereopsis – hence the phrase 'independent s.f.-tuned stereopsis channels'.

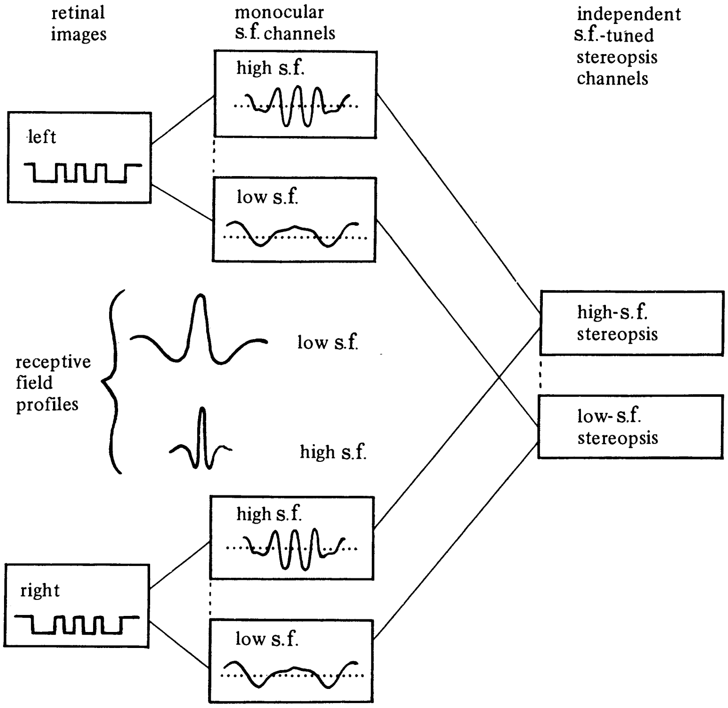


FIGURE 1. The independent s.f.-tuned channels model of stereopsis. The stereopsis boxes represent processes mediating both local and global stereopsis (see text for details). Reproduced from Frisby & Mayhew (1978*a*) by courtesy of Pion Limited.

The stereopsis model depicted in figure 1 does not specify what aspects of the s.f. filtered profiles are used for making left–right local matches. Frisby & Mayhew (1977*a*; see also Mayhew *et al.* 1977) implied a scheme whereby any 'left white point' could fuse with any 'right white point' (given the usual constraint imposed by Panum's fusional area); *mutatis mutandis* for left–right 'black point' matches. However, it would be equally possible and almost certainly preferable to use left–right peak and/or zero crossing matches (see later and also Marr & Poggio 1979). Thus the key feature of the model is simply that whatever local matches are made, they are effected between left–right convolution profiles separately for the different s.f. tuned channels, and with separate resolution of the ambiguity problem posed by false local fusions (i.e. separate s.f.-tuned global processes as well as local ones).

Given the independent channels model illustrated in figure 1, it becomes of interest to compare stereopsis contrast thresholds for 'complex' stereograms composed of multiple s.f. components, with similar thresholds for 'simple' stereograms composed of a single s.f. The independent channels model would predict that the threshold for stereopsis of a complex stimulus containing widely separated (two octaves different) s.f. components should be reached when the most sensitive channel reaches its own contrast threshold and that the presence of another s.f.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

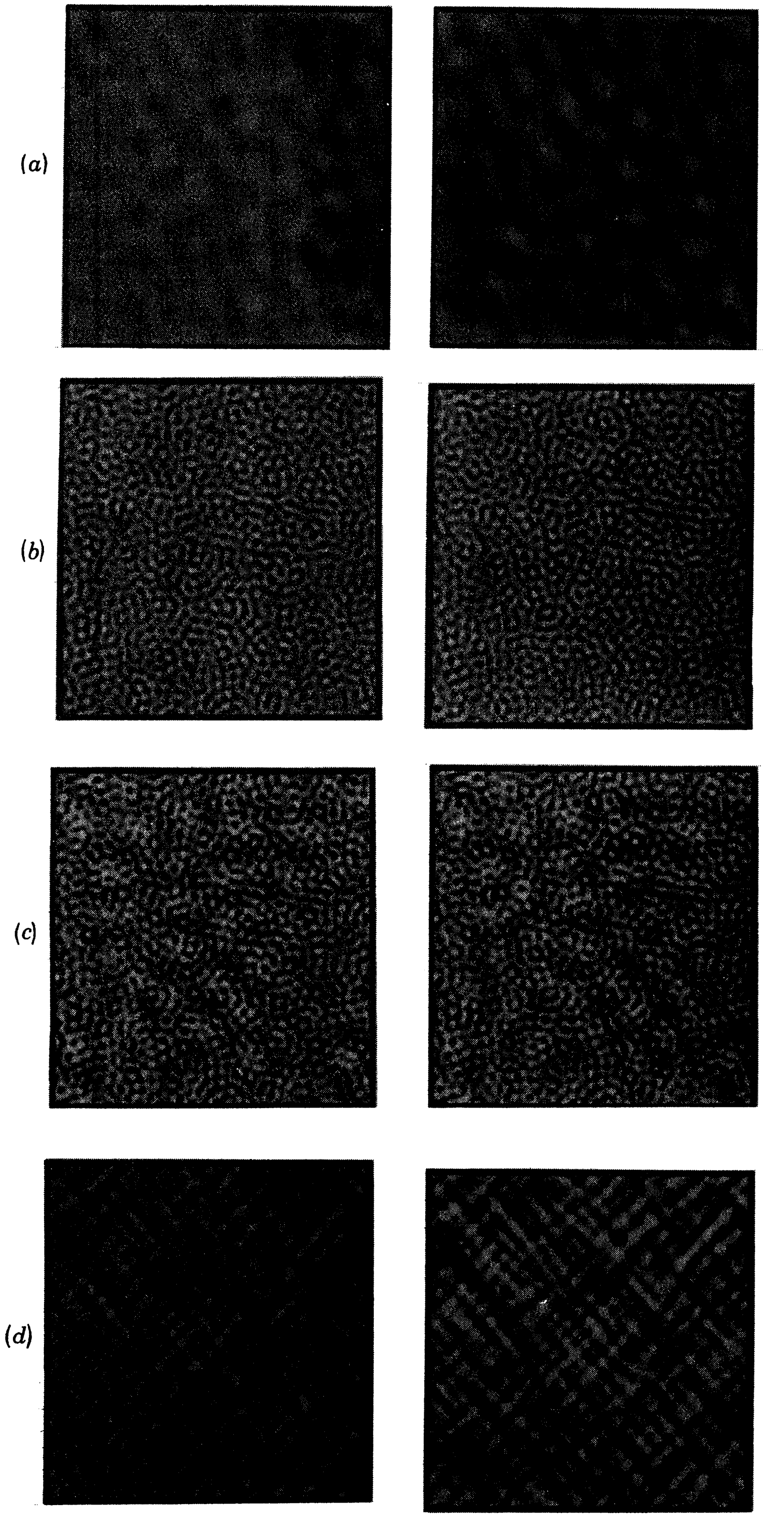


Figure 2. For legend see facing page.

component should not play an important role (given due allowance for probability summation). Mayhew & Frisby (1978*a*) reported several experiments testing this prediction, using both circularly filtered and orientated stimuli (figure 2). Contrast thresholds for stereopsis were always considerably lower than the independent s.f. channels model predicted and a variety of ways of saving the model from this falsification were tried out but found wanting. As a result, Mayhew & Frisby advanced the so-called SLUG model of stereopsis (several local unitary global) shown in figure 3, in which local matches remain s.f.-tuned (and thus capable of providing a locus for the Julesz–Miller masking effects) but with these feeding a single global stereopsis mechanism (cf. Marr & Poggio's $2\frac{1}{2}$ D sketch) for the resolution of local ambiguities and the build-up of surface descriptions.

At this point we asked: should SLUG be equipped with circularly symmetric local filters, or orientated ones, or both? Julesz & Miller (1975) used circular filtering for their demonstrations but the extensive psychophysical literature supporting the existence of orientated s.f channels (see, for example, Mostafavi & Sakrison 1976), when coupled with the finding that

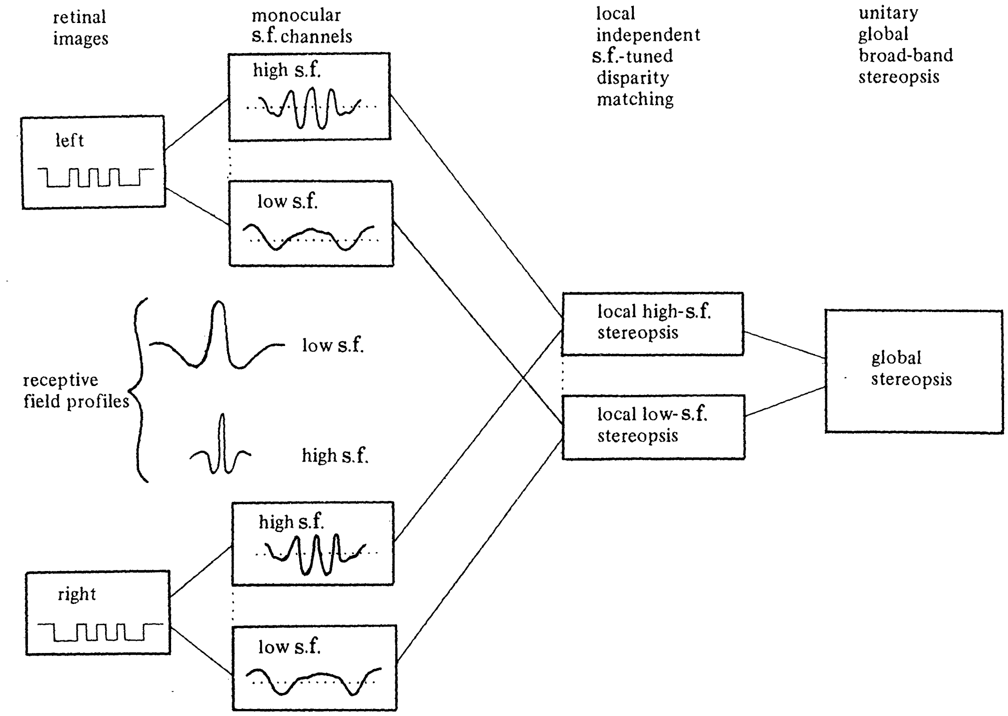


FIGURE 3. The SLUG model of stereopsis: Several local independent disparity-matching processes feed a Unitary Global stereopsis mechanism. Reproduced from Frisby & Mayhew (1978*a*) by courtesy of Pion Limited.

FIGURE 2. Contrast summation effects for stereopsis from complex stereograms containing two widely differing spectral components. (*a*) Low s.f. stereo pair (2.5 cycles/deg); (*b*) high s.f. stereo pair (10 cycles/deg); (*c*) complex s.f. stereo pair (2.5 and 10 cycles/deg combined in the contrast ratio of 1:3): this stereogram has a lower contrast threshold for stereopsis than predicted by the independent channels model of figure 1; (*d*) complex stereo pair made up of two oblique components: as for (*c*), stereopsis contrast threshold for this complex stimulus is lower than that predicted if its components were processed wholly independently for stereopsis. Note that these contrast summation effects do not hold for contrast thresholds for simple detection of these complex textures. All s.f. values hold (approximately) if the stereo pairs are viewed from about 10 × picture height. In all three stereo pairs, crossed eye fusion produces the percept of a central square floating above its surround. Reproduced from Mayhew & Frisby (1978*a*) by courtesy of Pion Limited.

cortical disparity units always seem to be orientated (see, for example, Poggio & Fischer 1977), encourages the belief in at least parallel provision of orientated disparity filters. We have conducted two psychophysical studies on this issue.

The first (Mayhew & Frisby 1978*b*) was an orientational equivalent of the Julesz–Miller masking study in which we discovered that if stereopsis from an orientated texture was masked by adding similarly orientated noise to one field, then rotation of the masking noise so that it would no longer interfere with orientated local matches did not succeed in releasing the stereopsis from the effects of the mask (figure 4). This is difficult to understand if local matches are made on orientated s.f. filtered profiles, although it might just be that the rotated noise, while freeing

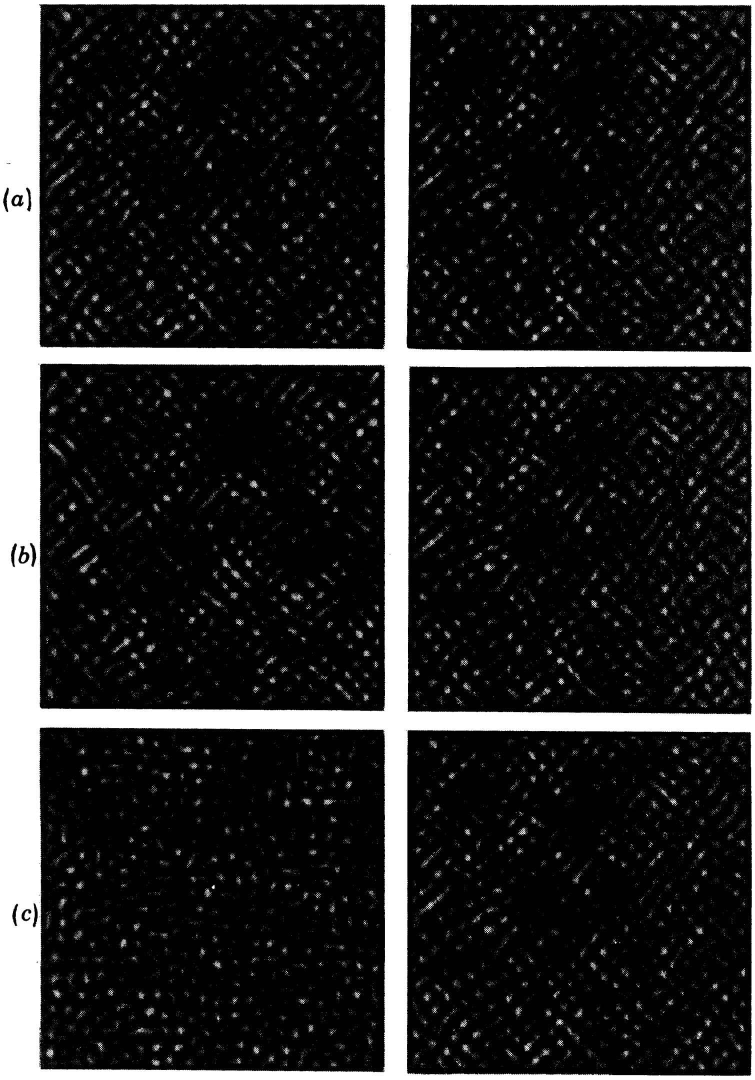


FIGURE 4. Stereopsis masking and orientational tuning. (*a*) Orientated-texture stereo pair: crossed-eye fusion produces the percept of a central square floating above its surround. (*b*) Same stereo pair as in (*a*) but masking noise of similar orientation and s.f. to that of the stereopsis signal has been added to the left half. Stereopsis is severely impaired and probably impossible for most observers. (*c*) Same stereo pair as in (*b*) but with the noise component of left field rotated by 45°. The quality of stereopsis is not improved by this rotation, even though the noise would now be stimulating different orientated s.f. channels from those triggered by the stereo signal. Reproduced from Mayhew & Frisby (1978*b*) by courtesy of Pion Limited.

the orientated matches, nevertheless produces interference at a global level from *ad hoc* spurious matches set up by non-orientationally tuned units. This possibility seems to us doubtful, however, in view of the known resistance of stereopsis to masking stimuli in many circumstances (see, for example, the many illustrations in Julesz (1971) to this effect).

Secondly, we have pointed out that orientated s.f. filters are in principle very poor devices for dealing with certain disparity cues, namely those from a surface with rapidly changing depths (Mayhew & Frisby 1979*a*). Consider, for example, the stereogram shown in figure 5 which depicts a series of horizontal corrugations (as though the observer were looking down on a corrugated roof). Vertically tuned s.f. filters could not *in principle* extract the depth information from this figure: they would inevitably have receptive fields spreading over several 'disparity rasters' and so smear the disparity cues hopelessly. It might be argued that the depth could be

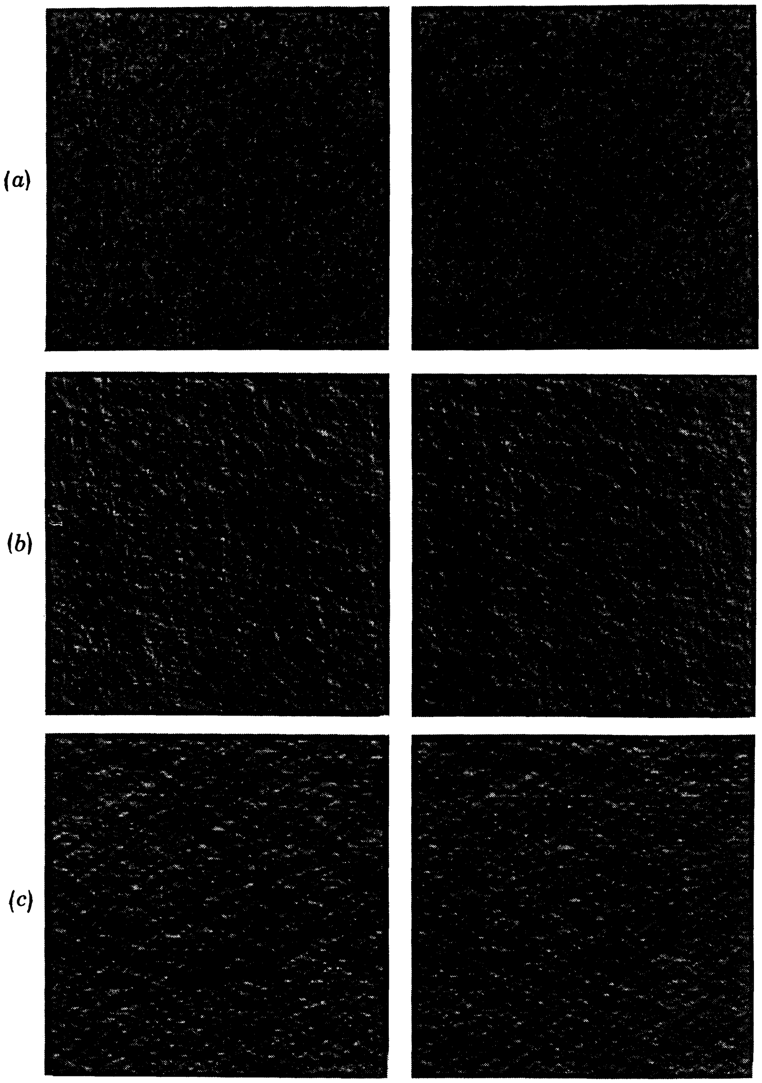


FIGURE 5. The processing of rapidly changing disparities is problematic for orientated s.f. channels. (*a*) Stereogram portraying a surface with near-horizontal corrugations. (*b*) Vertical filtering of (*a*) (orientation bandpass = vertical ± 45°) renders stereopsis impossible. (*c*) Horizontal filtering of (*a*) (orientation bandpass = horizontal ± 45°) severely impairs stereopsis.

mediated by horizontally tuned units in these circumstances but the poor quality of stereopsis deriving from a horizontally filtered version of the corrugated stereogram (figure 5*c*) suggests otherwise, as does the fact that many naïve subjects cannot obtain any depth whatsoever from this horizontally filtered stereo pair, whereas they can do so easily for the unfiltered original.

We conclude from these two studies that orientated s.f. filters do not seem to be used in the human visual system for establishing disparity matches. If valid, this conclusion would seem to demand a revision of Marr & Poggio's (1979) model of stereopsis which uses just such filters, although the revision required on this score may not be very fundamental (e.g. the simple substitution of circular for orientated filters may be enough). Of course, our rejection of orientated s.f. filters for stereopsis does not preclude other types of orientational selectivity embedded in the stereopsis mechanism. For a discussion of one alternative form of orientational tuning, we turn now to our computational model of stereopsis, called STEREOEDGE, which was designed with the foregoing psychophysical work in mind.

3. STEREOEDGE: A MODEL FOR THE COMPUTATION OF BINOCULAR EDGES

For STEREOEDGE (Mayhew & Frisby 1979*a*), a 'point' for stereo combination is an image location through which an image edge runs. To qualify for potential fusion, a pair of left and right edge points must possess roughly similar orientation and the same contrast polarity (e.g. a point that is part of a white-to-black edge in one field can fuse only with a white-to-black point in the other field, and not with one whose polarity is black-to-white). All possible local point-by-point combinations satisfying these requirements (and also that of falling within a realistically sized Panum's fusional area) are listed and a selection is then made of just those that conform to certain rules of figural grouping. In this way, STEREOEDGE takes advantage of the constraint that correct points for stereo fusion will be embedded in roughly similar contours in the two retinal images. Or course, these contours need not be object contours: they could be contours defining local elements within an object boundary and hence the constraint is applicable to a wide range of textures used for random-dot stereograms. Note that this constraint is quite different from either of those employed by Marr & Poggio (1976), i.e. the constraints of 'uniqueness of matches' (of suspect validity anyway, given Panum's limiting case) and 'depth continuity'.

By using rules of figural grouping for disambiguating local matches, the processing of disparity information by STEREOEDGE is intimately incorporated in those early visual computations contributing to figure-ground separation (Marr 1976). Thus STEREOEDGE represents an implementation of our earlier speculation that 'the processing of global disparity occurs in parallel with (rather than after) the very first stages of the computation of the symbolic descriptions of the visual scene . . ., the two processes sharing the same neural elements' (Mayhew & Frisby 1978*a*).

STEREOEDGE has two main parts, one monocular (called ZEROPOINT) and one binocular (called MATCH), and these will be described briefly in separate sections.

3.1. *ZEROPOINT: A procedure for finding monocular edge locations*

STEREOEDGE's first procedure, called ZEROPOINT, begins by convolving a grey level description of each monocular input image with a centre-surround operator (Laplacian). It then locates the points within each two-dimensional convolution profile through which contours pass. Left and right images are dealt with independently at this stage.

3.1.1. *The centre–surround convolution*

The design of the centre–surround operator can be varied at will so that it models a spatial frequency (s.f.) tuned channel of any desired characteristics. For example, figure 6*a* shows an input stereo pair portraying a teddy bear and figure 6*c* provides samples of the left and right convolution 'images' obtained by ZEROPOINT when its centre–surround operator is modelled approximately on the human contrast sensitivity function reported by Blakemore & Campbell (1969), i.e. when the centre–surround operator is equivalent to the broad-band channel shown in figure 7. Note that in a biological visual system, the whitish areas of figure 6*c* might be encoded in on-centre units and the blackish areas of off-centre units. Note also that in this particular case, the large areas of mid-grey do not signify zero convolution counts: the low-pass characteristics of the channel ensure that some response is made even to near-uniform areas of luminance (some retinal ganglion cells respond to overall level of luminance; Robson 1975). This feature has an advantage when ZEROPOINT proceeds to locate contours, as will become clear shortly.

Of course, given the work reviewed earlier implicating s.f.-tuned mechanisms in stereopsis, it is necessary to build in at some stage greater s.f.-selectivity than that of the broad-band channel of figure 7. This is an important issue to which we return later, having described STEREOEDGE's basic structure in terms of operations on a broad-band convolution profile. However, note that in view of our conclusion about the absence of orientated s.f.-tuned disparity processes, all convolutions used by STEREOEDGE are provided by circularly symmetric filters.

3.1.2. *Locating points on contours*

Contour boundaries in convolution images such as those of figure 6*b* are marked by positive–negative cross-overs (for figure 6*c*, white–black transitions), and ZEROPOINT is designed to obtain a description of where these zero crossings (z.cs) occur. Thus ZEROPOINT records the location of each z.c. (figure 6*d*) and ties to each one a description of:

(*a*) contour polarity (i.e. which side of the z.c. positive or negative);

(*b*) contour orientation (obtained by measuring the positive–negative gradient in eight orientations around each z.c. and then taking the contour orientation as that of the orientation providing the minimum gradient);

(*c*) contour gradient perpendicular to contour orientation.

Note that it is necessary to avoid logging positive-to-zero and negative-to-zero returns on either side of a contour boundary as 'genuine' z.cs: obviously they are not. In fact, such 'returns-to-zero' are infrequent for a broad-band operator with a low-pass characteristic of the kind shown in figure 7: usually, in such cases, convolution counts each side of a boundary stay either slightly positive or slightly negative, a fact that helps avoid logging 'spurious' z.cs. However, a few returns-to-zero are to be expected even in a broad-band channel and ZEROPOINT deals with them by eliminating any cross-overs that fall below a certain threshold gradient size. In narrow-band channels, returns-to-zero are much more common and also need a different approach for their elimination, the best one probably being to select as genuine z.cs only those cross-overs that occur in the same location in more than one channel (see later).

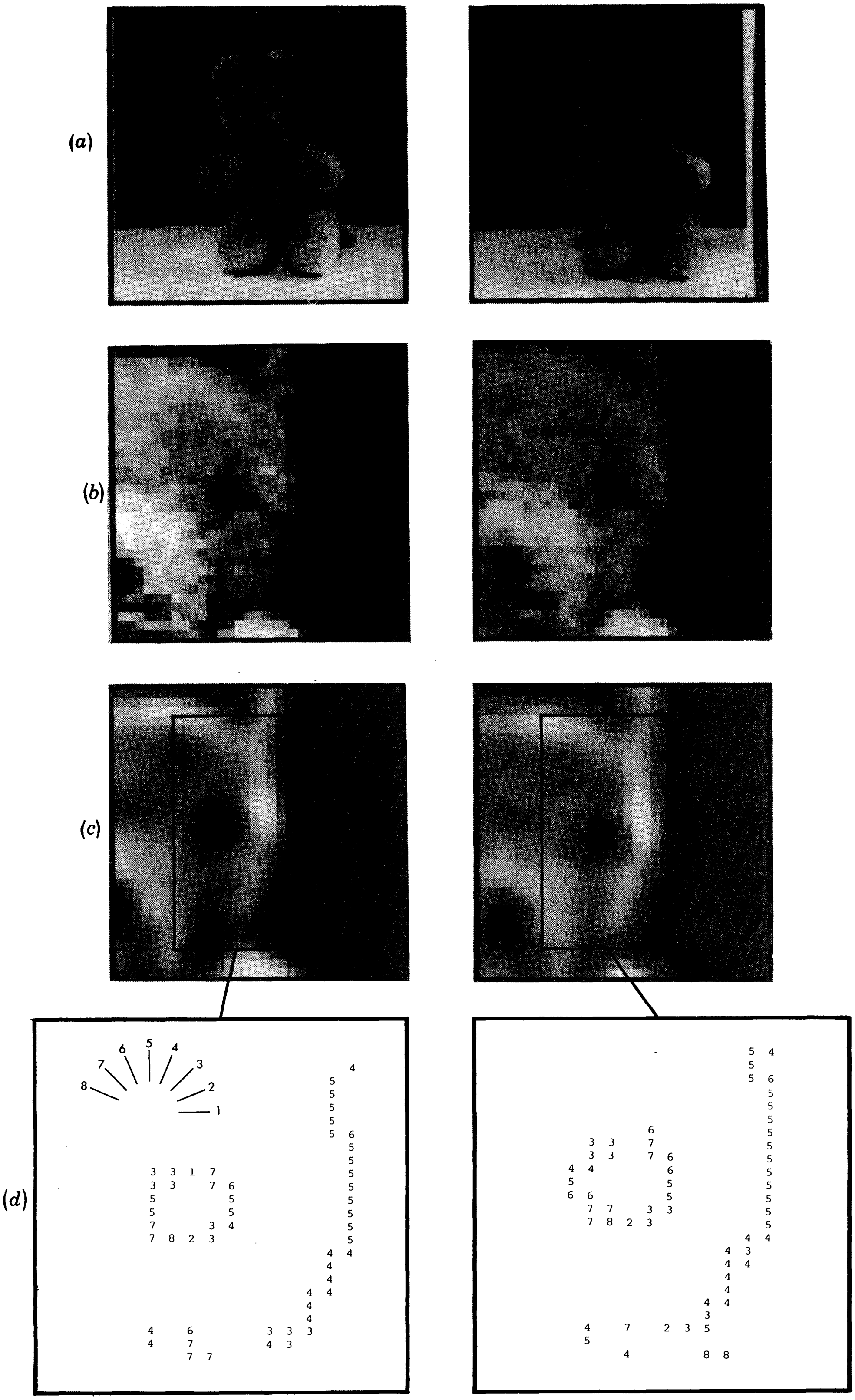


Figure 6 *a*–*d*. For legend see page 107.

3.2. *Match: A procedure for the binocular combination of edge locations*

Monocular 'points' used by MATCH for binocular combination are z.c. locations of similar contour polarity and orientation discovered by applying ZEROPOINT separately to each monocular image. However, similar left and right z.cs are combined (Panum's fusional area = ± 3 pixels horizontally, ± 1 pixel vertically) only if certain neighbourhood constraints are met around each potential z.c. fusion. The procedures applying these constraints embed certain principles of

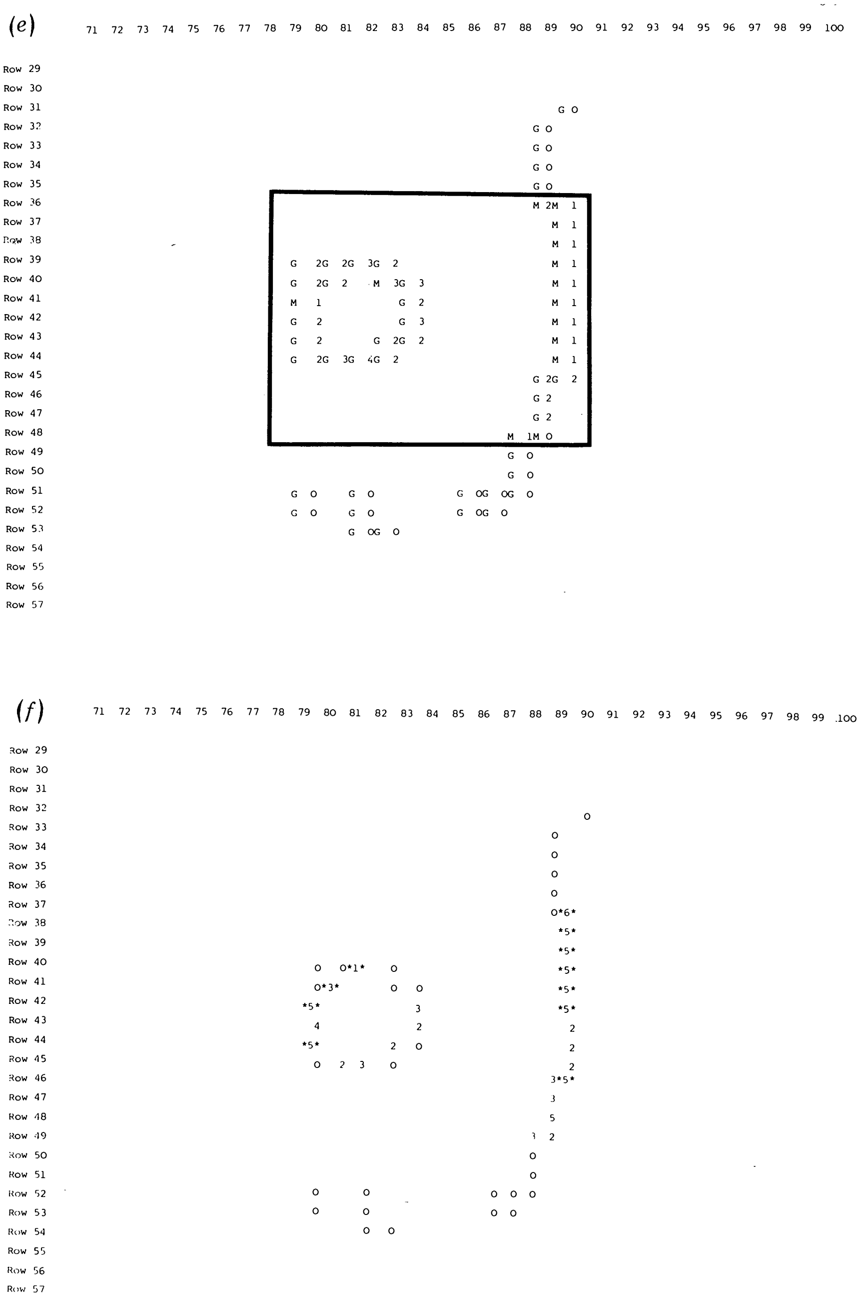
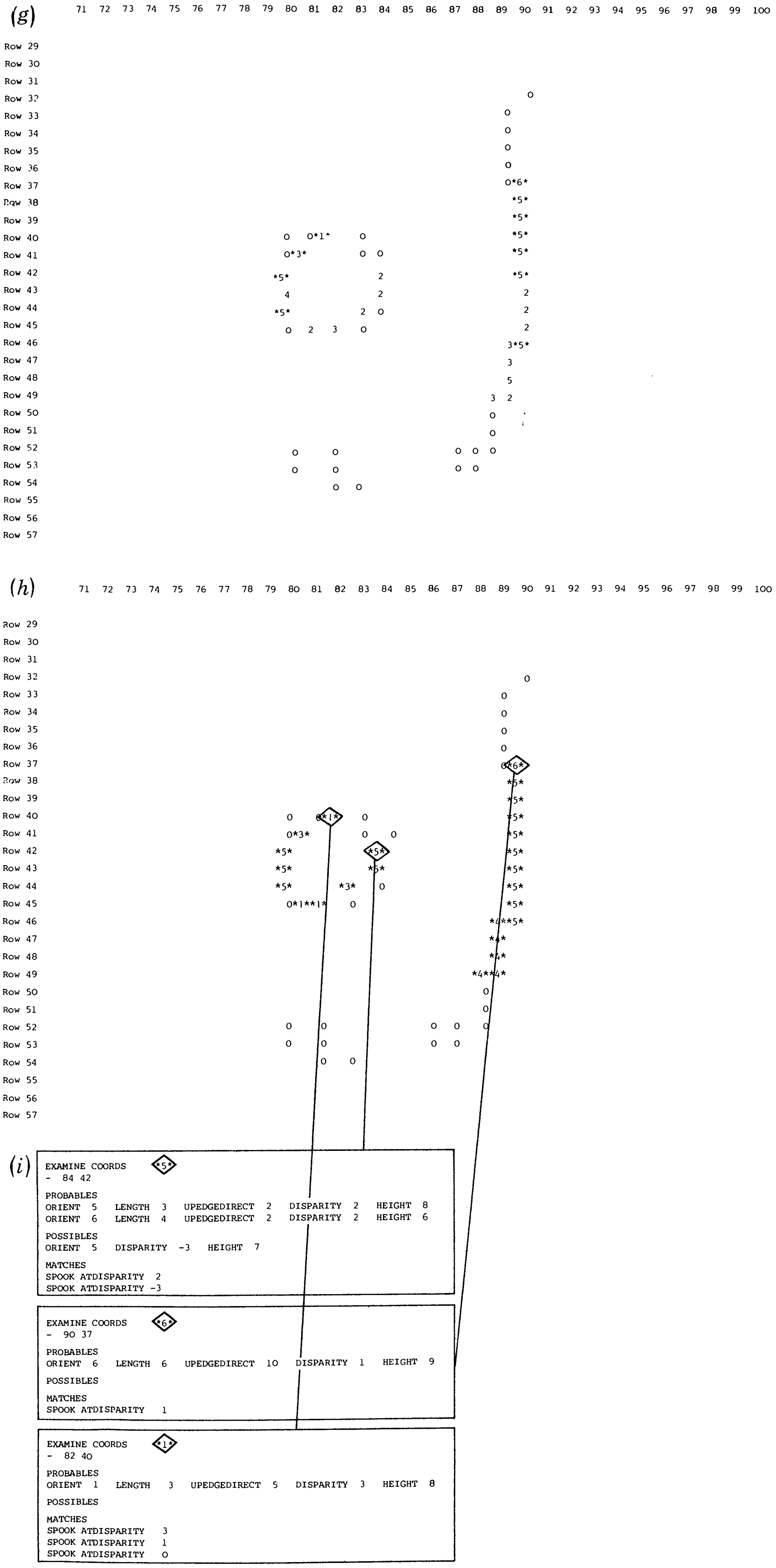


FIGURE 6*e–f*. For legend see page 107.



```
EXAMINE COORDS      *5*
-  84 42
PROBABLES
ORIENT  5   LENGTH  3   UPEDGEDIRECT  2   DISPARITY  2   HEIGHT  8
ORIENT  6   LENGTH  4   UPEDGEDIRECT  2   DISPARITY  2   HEIGHT  6
POSSIBLES
ORIENT  5   DISPARITY  -3   HEIGHT  7
MATCHES
SPOOK ATDISPARITY  2
SPOOK ATDISPARITY -3

EXAMINE COORDS      *6*
-  90 37
PROBABLES
ORIENT  6   LENGTH  6   UPEDGEDIRECT  10   DISPARITY  1   HEIGHT  9
POSSIBLES
MATCHES
SPOOK ATDISPARITY   1

EXAMINE COORDS      *1*
-  82 40
PROBABLES
ORIENT  1   LENGTH   3   UPEDGEDIRECT  5   DISPARITY  3   HEIGHT  8
POSSIBLES
MATCHES
SPOOK ATDISPARITY   3
SPOOK ATDISPARITY   1
SPOOK ATDISPARITY   0
```

FIGURE 6*g–i*. For legend see opposite.

figural grouping, and in this way ambiguity resolution among all possible local z.c. matches is achieved.

The strategy used by MATCH for constraining which z.c. combinations are finally utilized is to allow only these z.c. matches that are collinear (or quasi-collinear) in any particular depth plane. (In fact, we have explored extending the general scheme by allowing collinearity across depth planes, i.e. allowing collinear tilts, and there seem to be no difficulties in principle with this extension.) MATCH employs its collinear stratagem at various levels to select from a complete list of all possible z.c. matches (figure 6*e*) only those z.c. matches that reflect 'correct' edge fusions.

(*a*) First, a list is generated of all possible cyclopean assertions about short edge sections. Each such section is called a MICROEDGE and each one must be composed of at least three (quasi-) collinear matching z.cs in each eye's image spanning a distance of up to 7 pixels. Gaps of up to 2 pixels are allowed within each MICROEDGE and, as mentioned earlier, z.c. matches must share the same polarity and similar orientation (±45° seems satisfactory and matches

FIGURE 6. STEREOEDGE dealing with a stero pair of a natural scene (see text). (*a*) Grey level images (128 × 128 pixels) of the stereo halves so arranged that crossed-eyed fusion produces appropriate depth effects (i.e. right stereo half on the left hand side of figure). (*b*) Grey levels (32 × 32 pixels) in the region of the teddy bear's left eye and enlarged to illustrate details of STEREOEDGE's operation in subsequent figures. (*c*) Convolution outputs obtained from using the filter described in figure 7. (*d*) Zero crossing (z.c.) locations discovered in the right and left convolution outputs of (*c*) in the regions demarcated by the boxes. The numbers of each z.c. location indicate the orientation associated with that z.c. (number code for orientations inset in the left half). Note that equivalent right–left contours are represented by collections of z.cs that differ in their details, a fact that reflects faithfully the grey level images from whence they came. Any competent stereopsis processor must be able to deal with such left–right image differences. (*e*) Potential z.c. matches obtained by combining the two halves of (*d*). Computer storage limitations necessitate that a restricted region is dealt with at any one time, shown here by the inset window. The search for potential z.c. matches begins with the right eye (in a sense, therefore, treated as the 'dominant eye'). Thus for each z.c. in the right eye, a search is made along the same horizontal raster in the left eye (Panum's fusional area = ±3 pixels) for a suitable z.c. for matching. Requirements for a match are (i) same contrast polarity and (ii) roughly similar orientation (±45°). If no suitable left z.c. is found, then the vertical extent of Panum's area is relaxed to ±1 pixel of vertical disparity. If only one match is found, then this is coded with an M plus a number which represents the disparity of the match (±3 pixels disparity, equivalent to about ±8′ disparity if the stereo pair is viewed from about 10 × picture height. If more than one match is found, then this is shown with a G plus a number which gives the number of potential matches (G because the number includes ghosts). (*f*) Iteration 1 of the filtering algorithm. If a z.c. location is found to have just one possible match, then its orientation is given as a number code (see orientation code inset in (*d*)) and an asterisk is bound to it on either side. If a z.c. location has more than one possible match, then asterisks are absent and only a number is given which shows the number of possibilities. Z.c. locations with no possible matches according to the point reached by the algorithm at this stage are shown with a zero. (*g*) Iteration 3 of the filtering algorithm. Note that little change has occurred since iteration 1 for this particular region of the scene, the first iteration having done virtually all the work possible given the filtering constraints embedded in the algorithm. Coding as for (*f*). (*h*) Final 'best bet', achieved by resolving any ambiguity left over from the filtering algorithm by selecting the 'strongest' z.c. match at any ambiguous location, i.e. selecting that z.c. match with the largest weighting function. Note the successful resolution of ambiguities recorded in (*e*), (*f*) and (*g*). Coding as for (*f*). (*i*) Details of the data base provided by STEREOEDGE. Specimen print-outs are shown for just three of the z.c. locations given in (*h*). 'Probables' are a list of z.c. matches that survive right up to the 'best bet' stage; 'possibles' are those z.cs that have been excluded at some stage but might later be resuscitated and so are kept in store. The parameters tied to each entry are: ORIENT, orientation given as a number (code shown inset in (*d*)); LENGTH, the number of z.cs contributing to the MICROEDGE in which the z.c. in question is embedded; UPEDGEDIRECT, the direction of the maximum gradient at the z.c. expressed on a 16 point scale that gives information about contrast polarity; DISPARITY, ±3 pixels; WEIGHT, weighting function as described in text. The list of MATCHES records ghosts that have been eliminated. Note the successful location of the edges of the eye in depth planes in front of the edge of the head. Note also that although STEREOEDGE prints out individually its matched z.cs, it would be a trivial matter to group these into larger-scale edge assertions.

involving horizontal z.cs seem to cause few problems). Tied to each possible MICROEDGE assertion is the disparity and orientation of points along its length, and a weighting. The latter is a crude evaluation function: the longer and straighter the MICROEDGE the better; gaps in the MICROEDGE reduce the weight. Figure 6*f* illustrates the kind of output available at this stage.

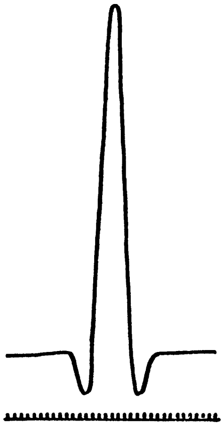


FIGURE 7. Receptive field profile used for producing the convolution images shown in figure 6*c*. The scale below the profile is in image pixels. The receptive field is roughly equivalent to a broad-band channel that would produce the contrast sensitivity function described by Blakemore & Campbell (1969).

(*b*) Next, the list of possible MICROEDGE assertions is pruned down by taking into account certain wider neighbourhood relationships. To stay in the list, a MICROEDGE must be either very strong (large weighting), or the only one in its vicinity, or else supported at one or other end by a neighbouring MICROEDGE that shares a similar orientation and disparity. (In a monocular curvilinear aggregator of the same general type, we have included the capability for allowing support at bends and corners to generate angle assertions. This extension is, however, too demanding of computer storage to be run in binocular mode on our machine at present.) The general elimination procedure is a parallel iterative Waltz-type filtering algorithm (Waltz 1975) in which MICROEDGE assertions without neighbourhood support (cf. incompatible blocks–world line labelling in Waltz 1975) are removed from the current list. MICROEDGES that are eliminated in this way, however, are not completely discarded but instead entered in a list of 'possibles'. Such MICROEDGES are no longer considered in successive iterations of the algorithm but they are recovered should the algorithm remove all MICROEDGE candidates in a given vicinity. (This latter eventuality often happens with z.cs towards the end of an edge, and particularly at the end of thin lines where the terminal z.c. has an orientation perpendicular to the z.cs forming the body of the line, so rendering it ineligible for support from these neighbours). The algorithm typically needs only three iterations to converge (figure 6*f*, *g*).

(*c*) As will be apparent from the foregoing, the constraints embedded in the curvilinear filtering algorithm often fail to force a unique interpretation, an outcome which contrasts with the blocks–world situation where the available constraints are so much stonger. Here, quantization fuzz spawns a blur of similar, mutually supporting and essentially equivalent MICROEDGE assertions so that it is necessary to have an additional step to determine which MICROEDGES are to be allowed to survive to the end. It turns out that it is sufficient simply to use the weighting function to select the 'best' of any competing group (figure 6*h*).

3.3. *The need for s.f. information*

For the purpose of computing primal sketch assertions, Marr (1976) used linear orientated s.f. channels as the measurement devices whose outputs were parsed both for descriptions of the orientations of contours in the input and for descriptions of contour type. S.f. information proved essential for the latter objective, e.g. for assertions of EDGES of various degrees of fuzziness. As will be clear from the foregoing account of STEREOEDGE, we have found it computationally convenient (as well as psychophysically sensible) to extract the orientation structure of an input image by using nonlinear orientated grouping processes operating upon non-orientated convolution profiles. We now turn to the question of whether our general scheme can be expanded to incorporate s.f. information essential to descriptions of edge type.

First, we note that as far as edges are concerned, s.f. channels with different tuning produce z.cs that all tend to fall in the same location (figure 8). As a result, z.cs in Laplacian convolution profiles seem ideal candidates for building up a good 'skeleton' around which to integrate

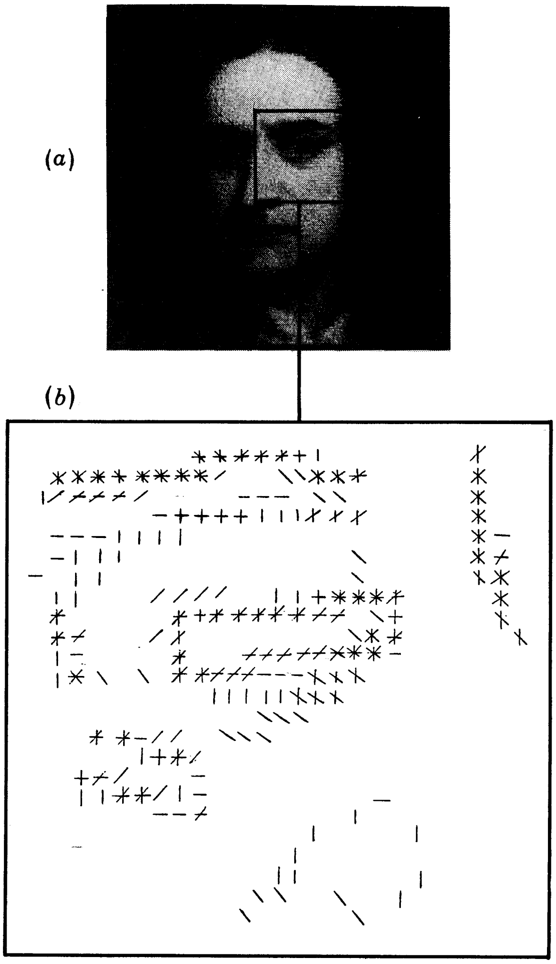


FIGURE 8. (*a*) Grey level input image (128 × 128 pixels) of Sir Isaac Newton. (*b*) Z.cs found by four differently tuned s.f. channels (all non-orientated) on a 32 × 32 pixel section of figure 8*a*: z.cs found by the 2.6 cycles/deg channel are shown with \ , by the 3.6 cycles/deg channel with |, by the 5.2 cycles/deg channel with —, and by the 7.2 cycles/deg channel with /. All s.fs apply for a viewing distance of figure 8*a* of about 10 × picture height. Note that z.cs in different channels show quite good alignment.

s.f.-tuned information relating to edge type. In contrast, peaks in Laplacian profiles from edges fall on either side of the edge's location, with their exact separation depending strongly on the s.f. tuning of the Laplacian and the slope of the edge with which it is dealing.

Secondly, note that the situation is the other way round for lines, where the *peaks* in the Laplacian profiles show convenient alignment, with the z.cs spread out on either side of the line centre. Consequently, a search for peaks showing similar locations in different channel outputs seem a sensible strategy for locating lines. In fact, we have already written a program that uses the nonlinear grouping of peaks to obtain MICROLINE assertions and it is planned to extend this monocular program to one capable of describing binocular lines. The difference between the peaks–z.cs situation for edges and lines follows from the fact that the Fourier components for an edge tend to be in sine phase whereas those for a line are in cosine phase.

Thirdly, the gradient of a z.c. is related to the contrast and frequency of the input. Indeed, because the gradient of a z.c. is directly proportional to input frequency, it is tempting to think that the human system contains an implicit compensation for the high frequency cut of the contrast sensitivity threshold function (Georgeson & Sullivan 1975).

Fourthly, Marr (1976) showed how the local spectrum of peaks delivered by a collection of different s.f.-tuned convolutions could be parsed into a description of edge type (i.e. fuzzy edge, sharp edge, etc.). The same objective can be attained in much the same way working not from peak information (as did Marr (1976)) but from the spectrum of z.c. gradients produced by s.f. channels for any given edge. Indeed, Marr now uses this approach himself (see his paper in this symposium.)

Summarizing the above points regarding edge descriptions and s.f. channels, it seems that a nonlinear orientated grouping of z.c. locations, taking into account z.c. gradients provided by several (say three or four) non-orientated s.f.-tuned channels, is a convenient and sufficient basis for computing edge descriptions – their location, orientation, type and contrast. Moreover, and most importantly for present purposes, the cyclopean grouping processes implemented in STEREOEDGE, for utilizing disparity information given by a broad-band channel, need little extension to be able to cope with the extra information provided by several monocular s.f.-tuned inputs. We are currently extending STEREOEDGE to explore the idea that the description of edge type takes place at one and the same time as the utilization of disparity information. In this way, STEREOEDGE will exhibit a degree of s.f. selectivity which, we expect, will model the Julesz–Miller stereopsis masking effects with which we began our programme of research.

4. LIMITATIONS: THE NEED FOR POST-CYCLOPEAN AND OTHER PROCESSES

4.1 *Higher order grouping processes*

The binocular edge descriptions returned by STEREOEDGE are local assertions characterizing the properties of an edge along its length (i.e. the orientations, polarities and disparities of its constituent points). Obviously, such descriptions can contribute substantially to the rich data base of low-level assertions needed by such processes as texture discrimination, region finding, large-scale curvilinear aggregation for object boundaries, etc. (Marr 1976). Indeed, we have implemented a program (called FRECKLES because it computes BLOB descriptions of various types) that applies higher-order grouping processes to the 'pointillist' output delivered by STEREOEDGE. This program will be described elsewhere (Mayhew & Frisby 1980).

4.2 *The problem of global stereopsis*

Mention was made earlier of the distinctions between *local stereopsis* (individual point-by-point matches) and *global stereopsis* (a resolution of the ambiguity existing within the pool of potential local point-for-point matches, i.e. a selection of correct local fusions at the expense of false local fusions or 'ghosts'). STEREOEDGE, however, reduces virtually to zero the scale of the global problem by insisting that potential local matches must be made only from points with identical polarity and roughly similar orientation. Given these restrictions, and a sensible (realistically small) Panum's fusional area, very few ghosts appear when potential local fusions are listed for the teddy bear stereogram (see figure 6*e*). To be sure, some ghosts do appear but these are usually produced by horizontal edge sections i.e. by intrinsically ambiguous regions of an image as far as depth from disparity is concerned, an ambiguity which can only be sorted out by later stages 'filling in' between correct depth locations assigned to edge ends (see § 4.3). Other ghosts can appear as a result of quantization fuzz and a certain sloppiness at corners due to the ± 1 pixel of vertical disparity which is allowed. Of course, it is of considerable interest to see that such ghosts as do appear in figure 6*e* are eliminated by the simple principle of figural grouping embedded in STEREOEDGE, but the key feature to note is that the global stereopsis problem as such hardly exists when potential point-for-point matches are restricted by a sensible definition of what constitutes a point-for-point match.

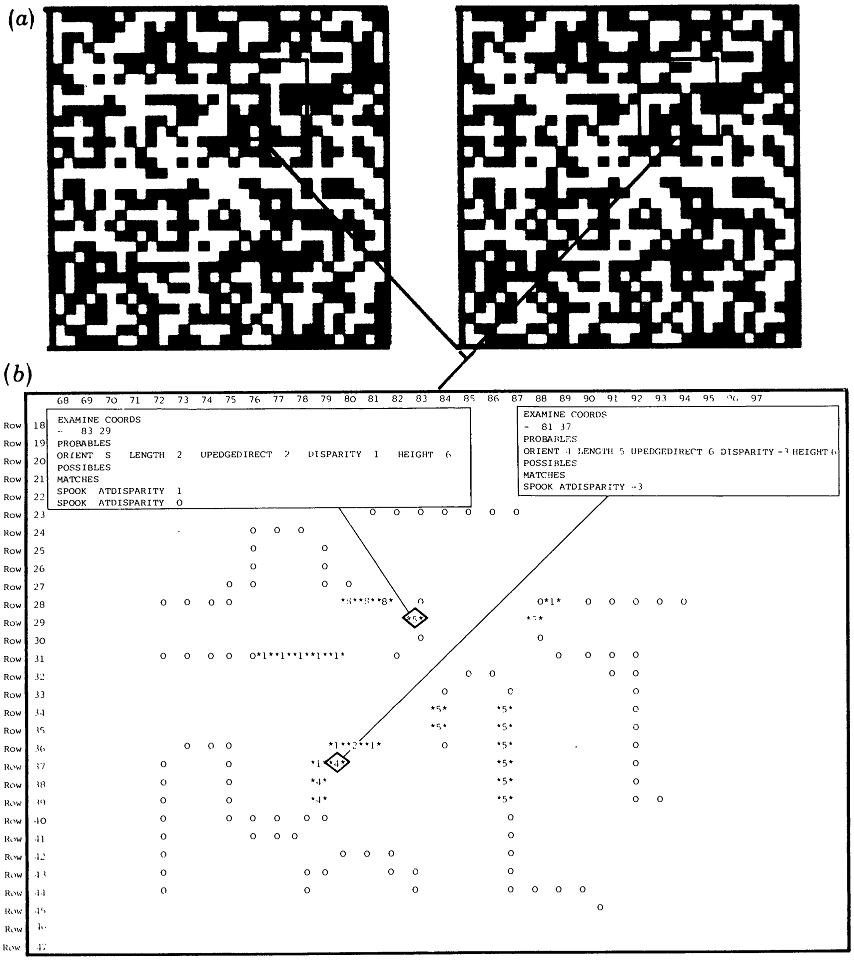
It might be asked, however, whether the situation would be as straightforward as this for a random-dot stereogram, which is normally thought to be a type of stereo input which presents a global problem in an especially acute form. The answer for a type of random-dot texture that suits STEREOEDGE's requirements for edges is given in figure 9. Somewhat surprisingly, ghosts are relatively rare here also and present no more problems than for the teddy bear stereogram. Of course, it might be that certain random-dot textures (such as those composed of myriad small dots) might not fall so readily to the present strategy, although it is interesting to observe that such textures often create severe difficulties for human subjects anyway and can produce very long stereopsis latencies (Frisby & Clatworthy 1975). Also, of course, stereo inputs can always be devised that create ghosts that are as numerous and strong as correct fusions (e.g. those with repeating sub-patterns). However, multiple stable states would then be just as characteristic of an artificial stereo system based on STEREOEDGE as they are of the visual system itself (*vide* the wallpaper illusion).

4.3. *The need for a* $2\frac{1}{2}$*D sketch*

STEREOEDGE delivers a description of edge locations and their relative depths, but it takes no account of the areas of image intervening between edges. Clearly, what is required is for the depth information tied to edges to be integrated with depth information tied to other image features (e.g. lines), and the whole used to build up a complete depth map of surfaces in the scene being viewed. In short, a computation of the kind envisaged by Marr for arriving at the $2\frac{1}{2}$D sketch seems the obvious next step (see Marr's paper in this symposium). Perhaps one component part of this computation could be 'filling in' the depth assigned to areas between edge locations, conceivably with the use of a minimum curvature algorithm of the type proposed by Ullman (1978) in connection with illusory contours.

Given that STEREOEDGE can fuse successfully patches of texture that are viewed with an appropriate vergence angle (i.e. the correct fusions falling within the allowable window defined by Panum's fusional area), a further requirement is to devise a mechanism that can shift the two eyes' inputs with respect to one another when zero or imperfect combination is discovered (cf. Marr & Poggio 1979). Given that the filtering algorithm used by MATCH measures neighbourhood support by way of selecting correct local fusions, it also seems possible in principle to use the output of this algorithm to drive 'vergence' movements if it records poorly matched inputs (i.e. weak neighbourhood support found). Thus the algorithm seems to lend itself naturally to a 'rivalry-driven' form of vergence movement control, the rivalry signal being terminated when computations embedded in the $2\frac{1}{2}$D sketch find a good fit between left and right descriptions for each part of the field of view. We are currently developing a stereo system of this kind.

A further factor to note as far as eye movement control is concerned is that we have recently



```
EXAMINE COORDS
-   83 29
PROBABLES
ORIENT  S   LENGTH  2   UPEDGEDIRECT  2   DISPARITY  1   HEIGHT  6
POSSIBLES
MATCHES
SPOOK  ATDISPARITY  1
SPOOK  ATDISPARITY  O
```

```
EXAMINE COORDS
-  81 37
PROBABLES
ORIENT 4 LENGTH 5 UPEDGEDIRECT 6 DISPARITY -3 HEIGHT 6
POSSIBLES
MATCHES
SPOOK ATDISPARITY -3
```

FIGURE 9. STEREOEDGE dealing with a random-dot stereo pair (see text). (*a*) Grey levels (128 × 128 pixels) of the stereo halves. Crossed-eye fusion produces a central protruding square of +4 pixels (roughly equivalent to a convergent disparity of 11′ for a viewing distance of about 10 × picture height). (*b*) Final output of STEREOEDGE: see figure 6*h* for details. Note the successful location in depth of the disparate square with respect to its surround. Thus one of the two z.cs that are picked out with full print-outs comes from the disparate square and has a −3 disparity (STEREOEDGE dealt with the stereo pair as though the square was receding, i.e. the left half of figure 6*a* was treated as the left stereo half). The other picked-out z.c. comes from the surround and has a disparity of +1 (because the stereo halves were shifted laterally with respect to one another by 1 pixel so that the 4 pixel disparity would fall within STEREOEDGE's ±3 pixel Panum's fusional area – a kind of 'eye movement' allowance). Thus the square – surround disparity difference is +1/−3 pixels disparity, so faithfully capturing the 4 pixel shift built into the random-dot stereogram.

discovered (Kidd *et al.* 1979) that texture contours (i.e. boundaries between image regions differing, say, in their orientational structure) can facilitate stereopsis and that they do so by guiding vergence eye movements. Accordingly, this kind of capability needs to be incorporated in a reasonably complete model of human stereopsis. Certainly any attempt to limit eye movement control to s.f. filters whose s.f. tuning determines their disparity selectively (Marr & Poggio 1979) seems doomed to failure in the light of our finding that texture contours which present no suitable s.f. signals of this kind can control vergence.

5. Conclusions

An important conclusion to be drawn from STEREOEDGE is that it is quite feasible to utilize disparity cues in the very early stages of the computation of edges. Indeed, the computation of binocular edges is no different in principle from the computation of monocular ones; it is simply somewhat easier because the use of disparity cues helps to distinguish various possibilities for joining up monocular edge points into higher structures. This conclusion nicely complements our psychophysical studies, referred to earlier, which led us to speculate that this is indeed the kind of computational strategy used by the visual system itself (Mayhew & Frisby 1978*a*).

Of greater interest, however, is the rather different perspective offered by our work for the concept of the orientationally selective 's.f.-tuned' channel. Instead of regarding these as providing a linear filtering stage of the kind used by Marr (1976), we find it convenient to view them as nonlinear grouping operators working to find orientated structures in the convolution profiles provided by several s.f.-tuned centre–surround channels. Marr has arrived at a similar conception independently (see his paper in this symposium) and we share his view that at least some of the units presently designated by as 'simple cells' by the neurophysiologists are in fact mediating the nonlinear grouping of zero crossings by way of representing contour assertions. Be that as it may, the concept of the 'orientation-tuned channel' as a nonlinear grouping operator is, we believe, consistent with a great deal of psychophysical evidence, particularly that on stereopsis as reviewed here, and at the same time it makes such channels the embodiment of an eminently sensible computational strategy.

We should like to thank Dr C. Brown for his help with computing equipment and Mr P. Stenton for the many hours of careful evaluation and testing that he has contributed to the 'debugging' of STEREOEDGE. The research was supported by Science Research Council Grant GR/A/50894.

References (Frisby & Mayhew)

Blakemore, C. & Campbell, F. W. 1969 On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *J. Physiol., Lond.* **203**, 237–260.

Frisby, J. P. & Clatworthy, J. L. 1975 Learning to see complex random-dot stereograms. *Perception* **4**, 173–178.

Frisby, J. P. & Mayhew, J. E. W. 1977*a* Global processes in stereopsis: some comments on Ramachandran and Nelson (1976). *Perception* **6**, 195–206.

Frisby, J. P. & Mayhew, J. E. W. 1978 The relationship between apparent depth and disparity in rivalrous-texture stereograms. *Perception* **7**, 661–678.

Georgeson, M. A. & Sullivan, G. D. 1975 Contrast constancy: deblurring in human vision by spatial frequency channels. *J. Physiol., Lond.* **252**, 627–656.

Julesz, B. 1960 Binocular depth perception of computer-generated patterns. *Bell Syst. Tech. J.* **39**, 1125–1162.

Julesz, B. 1971 *Foundations of cyclopean perception.* Chicago: University of Chicago Press.

Julesz, B. & Miller, J. 1975 Independent spatial-frequency-tuned channels in binocular fusion and rivalry *Perception* **4**, 125–143.

Julesz, B. & Oswald, H. P. 1979 Binocular utilisation of monocular cues that are undetectable monocularly. *Perception* **7**, 315–322.

Jung, R. 1973 *Handbook of Sensory Physiology*, vol. III, part iii A (ed. R. Jung), pp. 3–152. Berlin: Springer-Verlag.

Kidd, A. L., Frisby, J. P. & Mayhew, J. E. W. 1979 Texture contours can facilitate stereopsis by control of vergence movements. *Nature, Lond.* **280**, 829–832.

Marr, D. 1976 Early processing of visual information. *Phil. Trans. R. Soc. Lond.* B **275**, 483–524.

Marr, D. & Nishihara, H. K. 1978 Visual information processing: artificial intelligence and the sensorium of sight. *Technol. Rev.* **81**, 2–23.

Marr, D. & Poggio, T. 1976 Co-operative computation of stereo disparity. *Science, N.Y.* **194**, 283–287.

Marr, D. & Poggio, T. 1979 A computational theory of human stereo vision. *Proc. R. Soc. Lond.* B **204**, 301–328.

Mayhew, J. E. W. & Frisby, J. P. 1976 Rivalrous texture stereograms. *Nature, Lond.* **264**, 53–56.

Mayhew, J. E. W. & Frisby, J. P. 1978*a* Contrast summation effects and stereopsis. *Perception* **7**, 537–550.

Mayhew, J. E. W. & Frisby, J. P. 1978*b* Convergent disparity discriminations in narrowband filtered random-dot stereograms. *Vision Res.* **19**, 63–71.

Mayhew, J. E. W. & Frisby, J. P. 1978*c* Texture discrimination and fourier analysis in human vision. *Nature, Lond.* **275**, 438–439.

Mayhew, J. E. W. & Frisby, J. P. 1979*a* Surfaces with steep variations in depth pose difficulties for orientationally tuned disparity-filters. *Perception* **8**, 691–698.

Mayhew, J. E. W. & Frisby, J. P. 1979*b* The computation of binocular edges. *Perception* **9**. (In the press.)

Mayhew, J. E. W. & Frisby, J. P. 1980 Computational and psychophysical studies towards a theory of human stereopsis. *Artificial Intelligence* (submitted).

Mayhew, J. E. W., Frisby, J. P. & Gale, P. 1977 Computation of stereo disparity from rivalrous texture stereograms. *Perception* **6**, 207–208.

Mostafavi, M. & Sakrison, D. J. 1976 Structure and properties of a single channel in the human visual system. *Vision Res.* **16**, 957–968.

Poggio, G. F. & Fischer, B. 1977 Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J. Neurophysiol.* **40**, 1392–1405.

Ramachandran, V. S., Madhusudan Rao, V. & Vidyasagar, T. R. 1973 The role of contours in stereopsis. *Nature, Lond.* **242**, 412–414.

Robson, J. G. 1975 Receptive fields: neural representation of the spatial and intensive attributes of the visual image. In *Handbook of perception*, vol. 5: *Seeing* (ed. E. D. Corterette & M. P. Friedman), pp. 81–116. New York: Academic Press.

Saye, A. & Frisby, J. P. 1975 The role of monocularly conspicuous features in facilitating stereopsis from random-dot stereograms. *Perception* **4**, 159–171.

Ullman, S. 1976 Filling-in the gaps: the shape of subjective contours and a model for their generation. *A. I. Memo* no. 367 Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Waltz, D. 1975 Understanding line drawings of scenes with shadows. In *The Psychology of computer vision* (ed. P. H. Winston), pp. 19–91. New York: McGraw-Hill.

Wilson, H. R. & Bergen, J. R. B. 1979 A four mechanism model for threshold vision. *Vision Res.* **19**, 19–32.

Wilson, H. R. & Giese, S. C. 1977 Threshold visibility of frequency gradient patterns. *Vision Res.* **17**, 1177–1190

Discussion

P. E. King-Smith (*Ophthalmic Optics Department, U.M.I.S.T., P.O. Box* 88, *Manchester M*60 1*QD, U.K.*). The authors demonstrate that there is no instantaneous perception of orientation in a pattern derived from superposing two narrow-band, perpendicularly filtered images of a random dot pattern, and suggest that this indicates that the output from orientation specific channels is not available for instaneous perception. Could an alternative explanation be that the bright and dark 'blobs' formed by superposing the two filtered images may stimulate *non-orientated* channels which, in turn, inhibit orientation-specific channels?

J. P. Frisby and J. E. W. Mayhew. The key question about orientated and non-orientated channels as far as we are concerned is: what are they doing? Until they have been assigned some functional role in a computational theory (see Marr's contribution to this symposium), it is difficult to evaluate how plausible any *ad hoc* putative inhibitory interactions between channels might be. The important point about our texture discrimination demonstrations is that they seem difficult to reconcile with any straightforward Fourier analysis theory of low-level visual function for which the linear orientated spatial frequency tuned channel was supposed to provide the physiological underpinning.

D. Marr (*Artificial Intelligence Laboratory, M.I.T., Cambridge, Massachusetts* 02139, *U.S.A.*). The idea that collinear grouping ('curvilinear aggregation'; Marr 1976) may be used to help solve the matching problem in human stereopsis is an interesting one. It is, however, not necessary to involve grouping at this stage (Marr & Poggio 1979). If an ambiguous match arises, the two candidates will almost always have disparities of opposite signs, and to remove ambiguity it is necessary only to consult the *signs* of neighbouring, unambiguous matches. In other respects, Frisby & Mayhew's program is equivalent to Marr & Poggio's theory, recently implemented by Grimson & Marr (1979).

Reference

Grimson, W. E. L. & Marr, D. 1979 A computer implementation of a theory of human stereo vision. Image Understanding Workshop, April 1979, A.I.lab., M.I.T., Cambridge, Massachusetts.

J. P. Frisby and J. E. W. Mayhew. We had not seen a report of Grimson & Marr's program, but now that we have we agree that there are some similarities between that implementation of Marr & Poggio's theory and our own program. Despite some common features, however, the differences seem to us to be more fundamental and important. Our program has been guided by psychophysical results suggesting that disparity processing is intimately integrated with the computation of symbolic descriptions of edges, blobs, lines, etc. Marr & Poggio, on the other hand, regard the extraction of disparity information as taking place within its own separate visual processing module. These different starting points have led to certain key differences between the programs in their use of orientated processes and in the role that they give to spatial frequency tuned channels.

Considering the orientation domain first, it is an important aspect of our approach that the computation of disparity and the extraction of local orientation structures are mediated by interrelated and mutually supporting processes. Our psychophysical results, however, indicated that orientated filters were not involved and so we turned to using orientated collinear grouping processes to guide the removal of ambiguity of zero crossings discovered in circularly filtered left–right images. Stereoedge was written, therefore, as a computational test of the idea that a nonlinear grouping process could have the dual function of extracting orientation structure and disparity information, and at the same time resolve any ambiguous matches that might arise. Interestingly, and significantly as far as the question of general approach is concerned, Grimson & Marr also chose to use circular filters rather than the orientated filters employed originally by Marr & Poggio, but they did so for the elegant computational reasons described by Marr & Hildreth (see Marr's paper in this symposium) rather than to make their program more psychophysically plausible.

Of course, we agree with Dr Marr when he says that Grimson & Marr's program demonstrates that it is not necessary to solve the stereopsis ambiguity problem by building in collinear grouping processes – but this is so at the expense of the strong assumptions made by Marr & Poggio about the relation between disparity processing and spatial frequency tuned channels. Thus Marr & Poggio develop the idea that disparities of different magnitudes are processed independently by different spatial frequency channels. Their theory holds that large disparities are dealt with by low spatial frequency units, with vergence movements initiated to bring into correspondence high spatial frequency channels dealing with small disparities. By limiting the size of the disparity that can be processed by any one channel to around plus or minus the width of the central region of the receptive fields of the units composing the channel, the theory avoids the problem of having to choose between many ambiguous matches. The key question here, however, is whether wholly independent disparity processing within different spatial frequency tuned channels actually takes place within the human visual system. The psychophysical evidence on this point is not yet clear and although we have contributed to the development of this kind of proposal ourselves (Mayhew & Frisby 1976; Frisby & Mayhew 1977*a*), our more recent work has found little support for it (Frisby & Mayhew 1977*b*; Mayhew & Frisby 1979*c*). Consequently, in our own theory we do not ascribe to spatial frequency tuned channels the role accorded to them by Marr & Poggio. Rather, as we note in our paper, it is a natural development of our program to take advantage of the contraints and correspondences between spatial frequency channel measurements that can be used for the computation of primitive assertions (Marr 1976), to force unique disparity assignments in most cases. Thus in our development of STEREOEDGE, local disparity matches will remain s.f.-tuned (as in a SLUG-type model: see our paper) but global processing to resolve ambiguities will utilize cross-s.f.-channel correspondences. Moreover, by taking advantage of these correspondences, together with those referred to above within the orientation domain, a much greater disparity range for local s.f.-tuned matches will be possible than in Marr & Poggio's model. Pilot work to date suggests that this approach is promising and it certainly fits in with our overall conceptual framework, namely to embed disparity processing intimately within the very first stages of the computation of primitive symbolic descriptions of edges, blobs and lines.

References

Frisby, J. P. & Mayhew, J. E. W. 1977*b* Spatial frequency tuned channels and stereopsis. Paper delivered to the Experimental Psychology Society, March 1977, University of Sheffield.

Mayhew, J. E. W. & Frisby, J. P. 1979*c* Convergent disparity discriminations in narrow-band-filtered random-dot stereograms. *Vision Res.* **19**, 63–71.

K. H. Ruddock (*Biophysics Section, Physics Department, Imperial College, London SW7 2BZ, U.K.*). In the authors' analysis of early visual processing, they postulate Laplacian operators, with non-orientated spatial response characteristics, that feed into orientation selective response units. Experiments with patterns containing two-dimensional spatial structure show that there are two classes of adaptation mechanism, one sensitive to symmetric 'spot-shaped' targets and the other to elongated, 'bar-shaped' targets (Naghshineh & Ruddock 1978). The former are monocularly and the latter binocularly driven, as is the case in the Frisby–Mayhew model. Both the adaptation data and independent threshold detection data (Burton 1976) indicate, however, that the non-orientation selective units operate both in parallel and in series with the orientation selective units. Have the authors considered the possibility of parallel operation of their two classes of response unit, perhaps as a means of incorporating additional stimulus characteristics such as colour and movement?

References

Burton, G. J. 1976 Visual detection of patterns periodic in two-dimensions. *Vision Res.* **16**, 991–998.

Naghshineh, S & Ruddock, K. H. 1978 Properties of length-selective adaptation mechanisms in human vision. *Biol. Cybernet.* **31**, 37–47.

J. P. Frisby and J. E. W. Mayhew. The short answer is no. While it might well be that other processes acting on the outputs of 'our units' might make explicit additional stimulus characteristics, our present concern is simply to develop a model capable of dealing with the early processing of monochromatic static three-dimensional scenes.

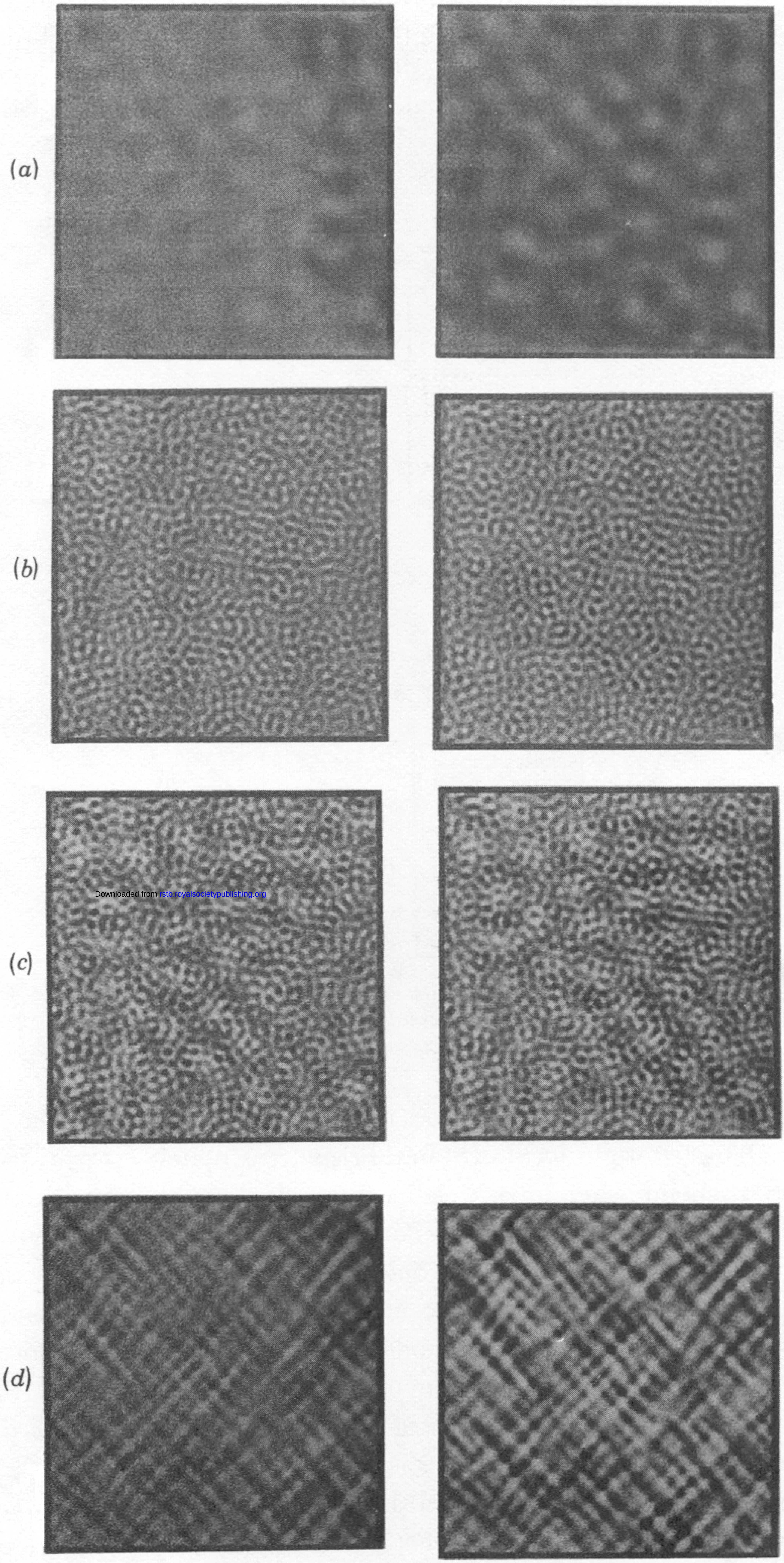


FIGURE 2. For legend see facing page.

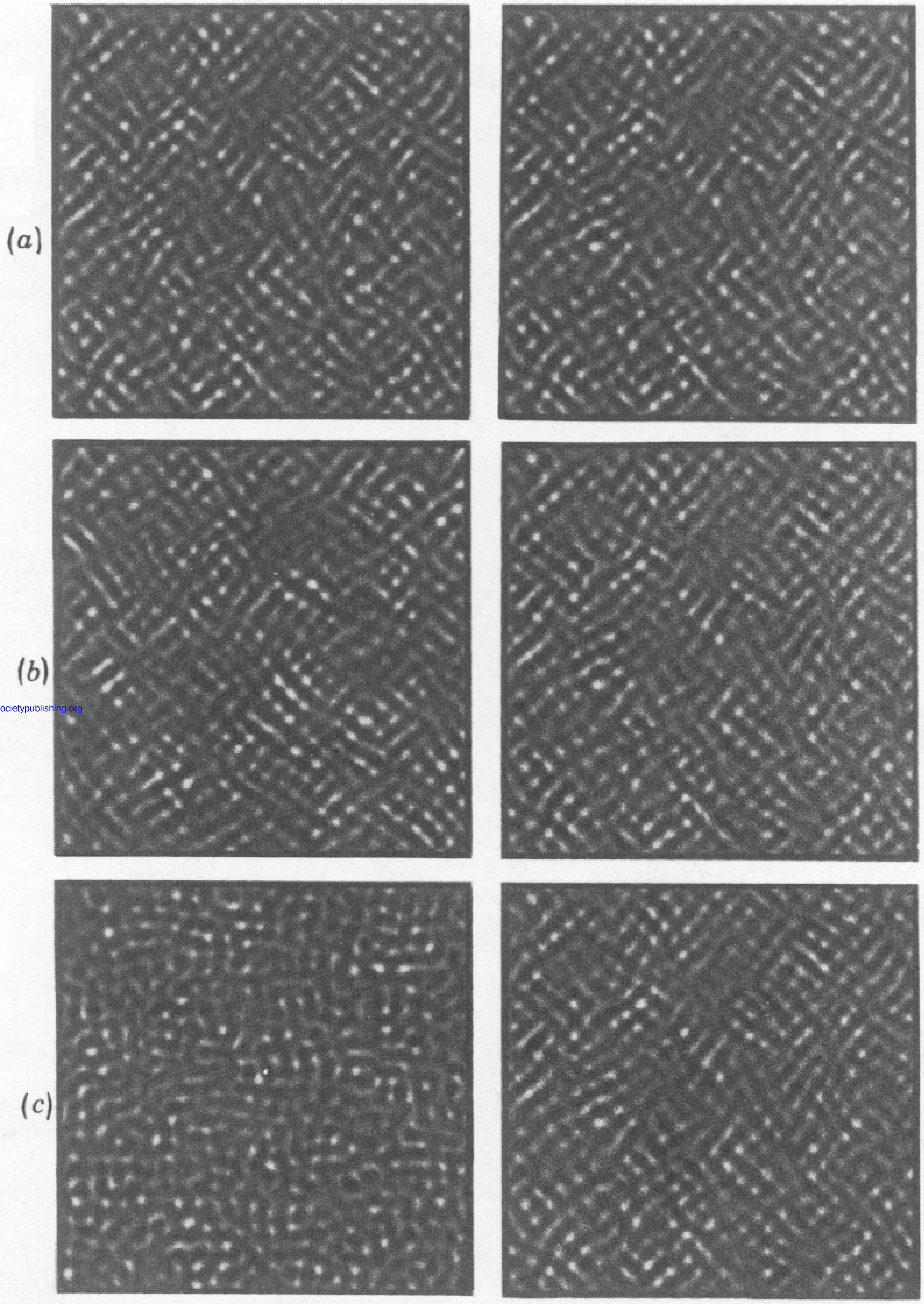


Figure 4. Stereopsis masking and orientational tuning. (*a*) Orientated-texture stereo pair: crossed-eye fusion produces the percept of a central square floating above its surround. (*b*) Same stereo pair as in (*a*) but masking noise of similar orientation and s.f. to that of the stereopsis signal has been added to the left half. Stereopsis is severely impaired and probably impossible for most observers. (*c*) Same stereo pair as in (*b*) but with the noise component of left field rotated by 45°. The quality of stereopsis is not improved by this rotation, even though the noise would now be stimulating different orientated s.f. channels from those triggered by the stereo signal. Reproduced from Mayhew & Frisby (1978*b*) by courtesy of Pion Limited.

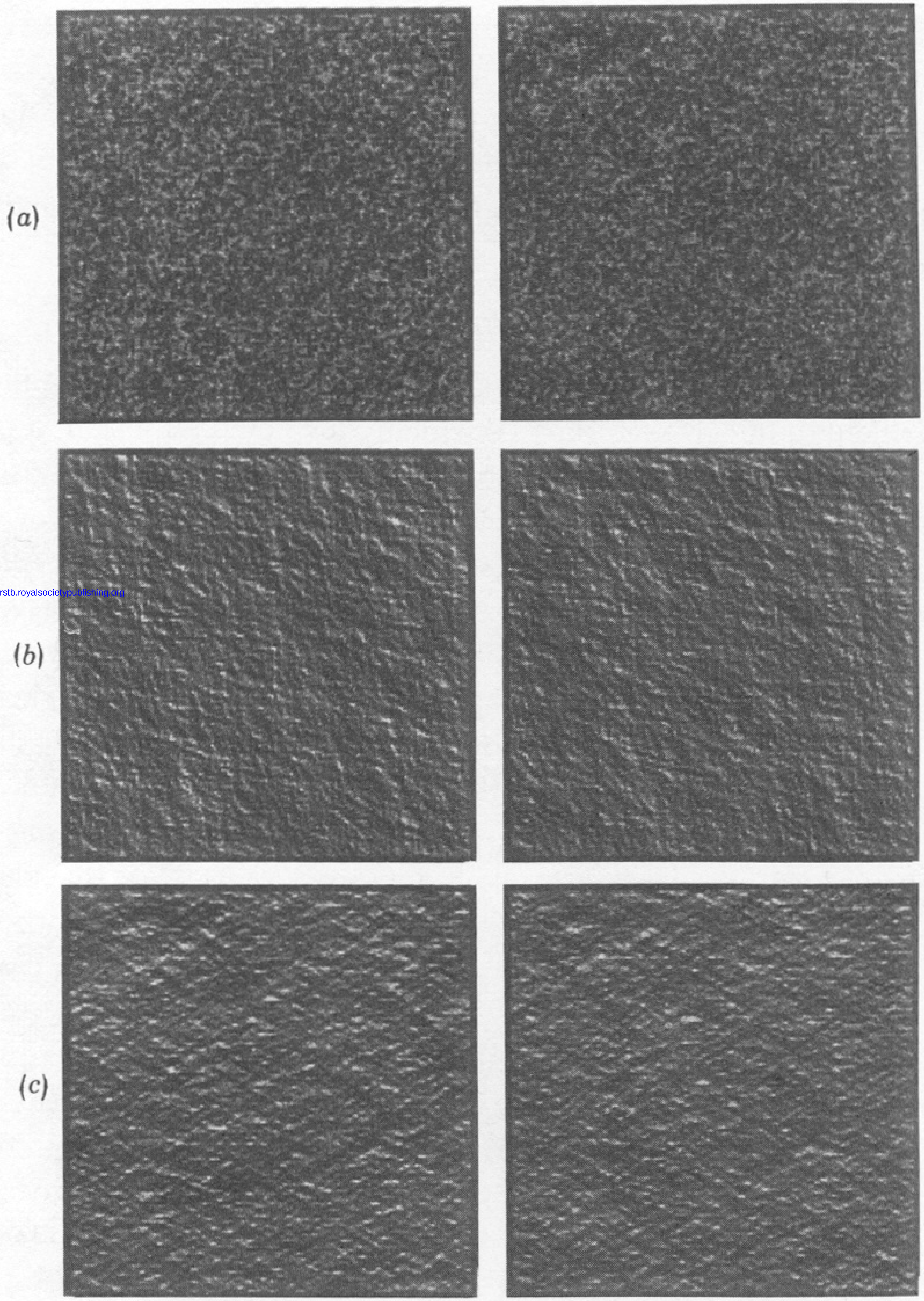


FIGURE 5. The processing of rapidly changing disparities is problematic for orientated s.f. channels. (*a*) Stereogram portraying a surface with near-horizontal corrugations. (*b*) Vertical filtering of (*a*) (orientation bandpass = vertical ± 45°) renders stereopsis impossible. (*c*) Horizontal filtering of (*a*) (orientation bandpass = horizontal ± 45°) severely impairs stereopsis.

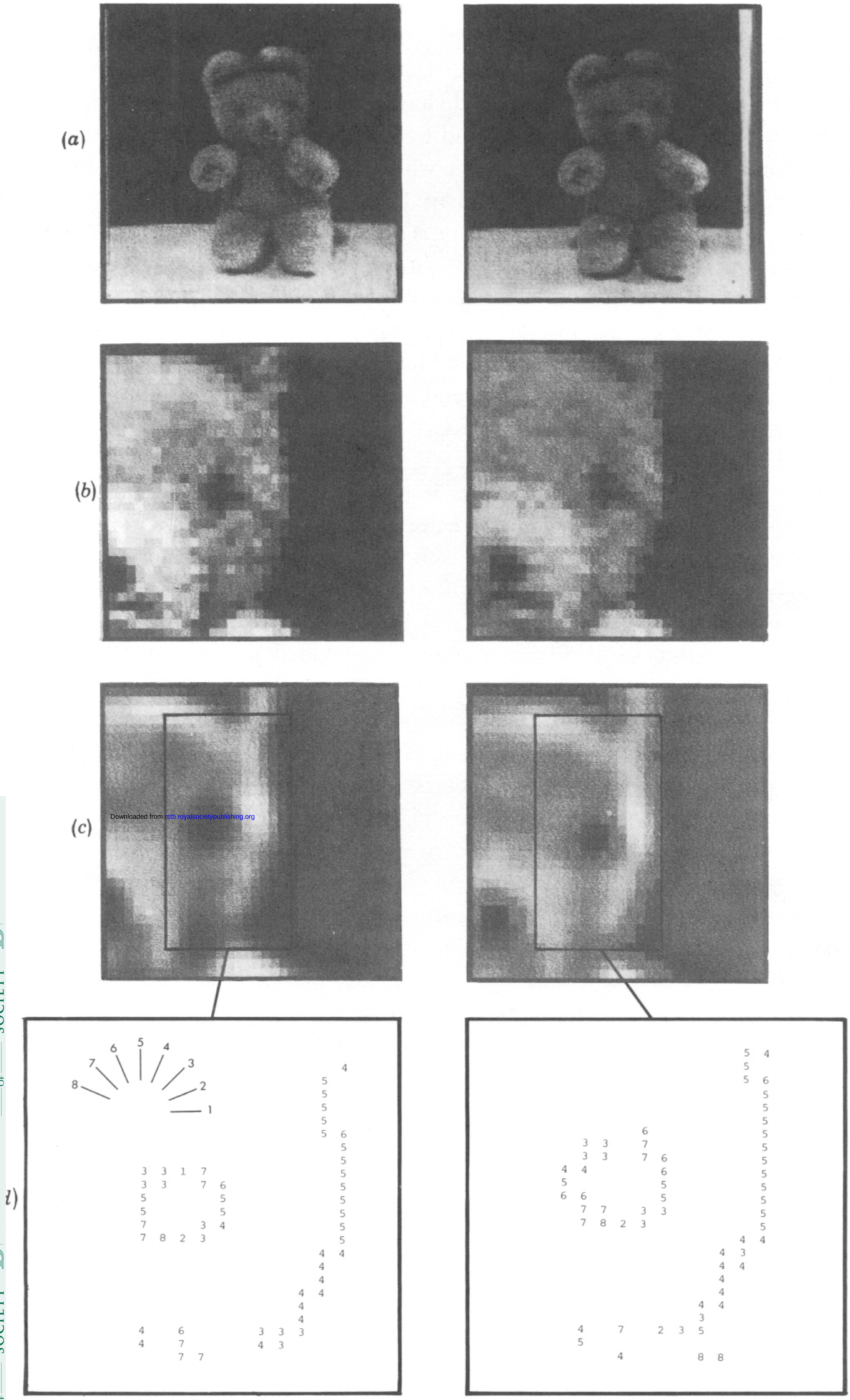


FIGURE 6 *a–d*. For legend see page 107.

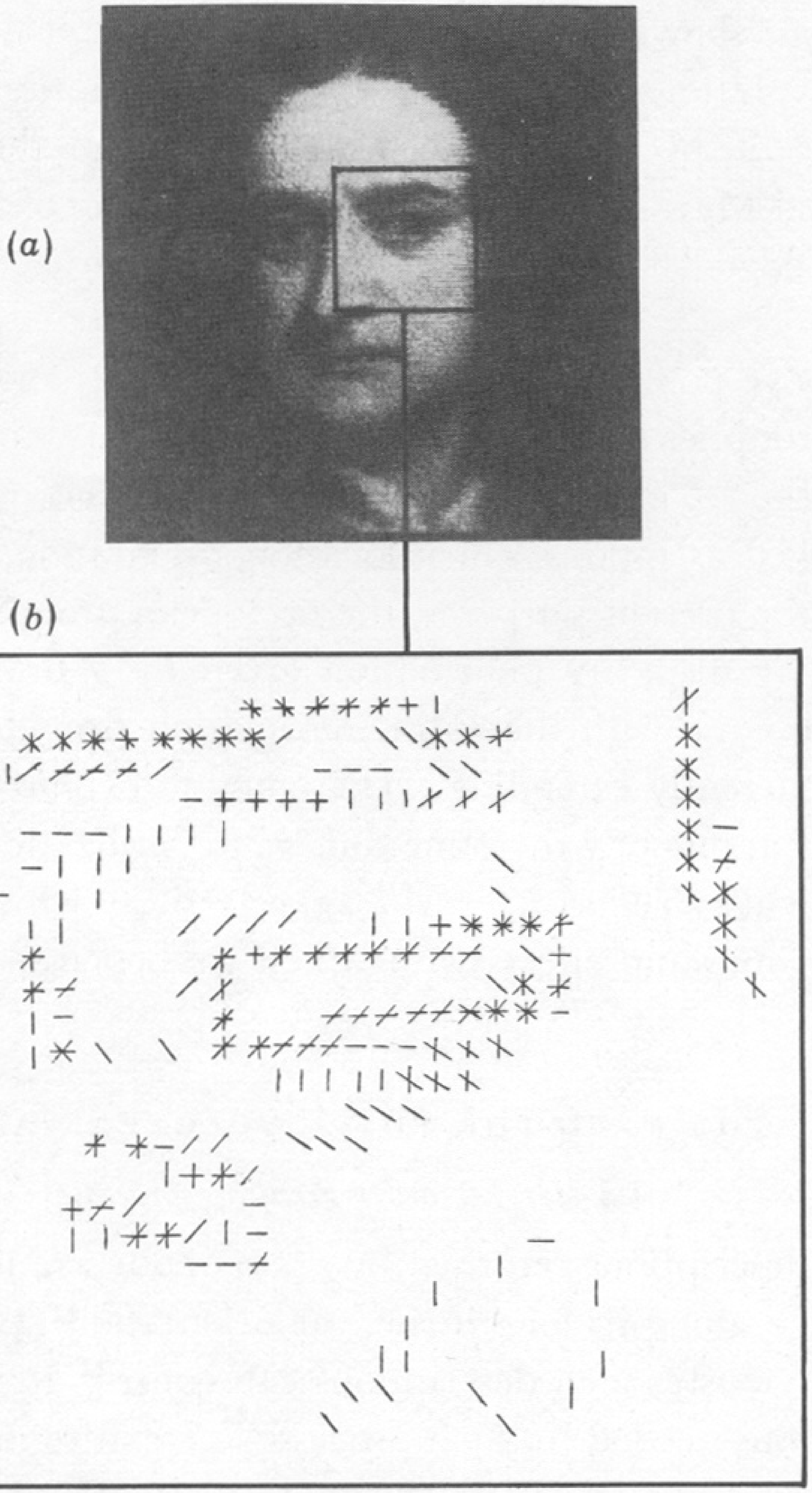


FIGURE 8. (*a*) Grey level input image (128 × 128 pixels) of Sir Isaac Newton. (*b*) Z.cs found by four differently tuned s.f. channels (all non-orientated) on a 32 × 32 pixel section of figure 8*a*: z.cs found by the 2.6 cycles/deg channel are shown with \, by the 3.6 cycles/deg channel with |, by the 5.2 cycles/deg channel with —, and by the 7.2 cycles/deg channel with /. All s.fs apply for a viewing distance of figure 8*a* of about 10 × picture height. Note that z.cs in different channels show quite good alignment.