Edges and bars: where do people see features in 1-D images?

Gillian S. Hesse, Mark A. Georgeson *

Neurosciences Research Institute, Aston University, Birmingham B4 7ET, UK

Abstract

There have been two main approaches to feature detection in human and computer vision—based either on the luminance distribution and its spatial derivatives, or on the spatial distribution of local contrast energy. Thus, bars and edges might arise from peaks of luminance and luminance gradient respectively, or bars and edges might be found at peaks of local energy, where local phases are aligned across spatial frequency. This basic issue of definition is important because it guides more detailed models and interpretations of early vision. Which approach better describes the perceived positions of features in images? We used the class of 1-D images defined by Morrone and Burr in which the amplitude spectrum is that of a (partially blurred) square-wave and all Fourier components have a common phase. Observers used a cursor to mark where bars and edges were seen for different test phases (Experiment 1) or judged the spatial alignment of contours that had different phases (e.g. 0° and 45°; Experiment 2). The feature positions defined by both tasks shifted systematically to the left or right according to the sign of the phase offset, increasing with the degree of blur. These shifts were well predicted by the location of luminance peaks (bars) and gradient peaks (edges), but not by energy peaks which (by design) predicted no shift at all. These results encourage models based on a Gaussian-derivative framework, but do not support the idea that human vision uses points of phase alignment to find local, first-order features. Nevertheless, we argue that both approaches are presently incomplete and a better understanding of early vision may combine insights from both.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Human vision; Psychophysics; Features; Edge detection; Derivatives; Local energy; Mach Bands; Contour alignment; Gaussian derivatives

1. Introduction

1.1. Luminance contours and their definition

For more than four decades since the pioneering studies of Hubel and Wiesel (1959, 1962) vision researchers have sought to understand the relationship between the response characteristics of visual cortical cells and the perceived structure of images. It is widely accepted that much of the key information in images lies in the spatial structure of contours, and it is now clear that visible contours can be defined by spatial transi-tions in many different kinds of image property, such as luminance and chromaticity (the so-called first-order properties) as well as higher-order properties such as local contrast, texture, motion and binocular disparity (for references and review see Regan (2000)). Our interest in this paper lies in what defines first-order, luminance contours for human observers.

Knowing what aspects of luminance variations in images gave rise to perceived contours would usefully constrain our computational models and our interpretation of physiological findings. Recent computational analyses have confirmed the importance of local, oriented edge- and bar-like structures in natural images (Bell & Sejnowski, 1997; Field & Brady, 1997; Olshausen & Field, 1996; van Hateren & van der Schaaf,

---

* Corresponding author. Tel.: +44 121 204 4119.
E-mail address: m.a.georgeson@aston.ac.uk (M.A. Georgeson).

1998) and have studied how perceptual grouping rules are related to the statistics of contours in natural images (Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001). Importantly, and despite earlier claims to the contrary, Elder (1999) showed that sparse edge maps can contain all the information needed for a fairly faithful re-construction of the original image. He thus showed that the edge map can be a 'perceptually lossless' representation of the image—provided that the local edge descriptions carry information about location, orientation, contrast and blur.

Surprisingly, however, with respect to human vision there is still much uncertainty over the most appropriate psychophysical definition for simple luminance contours such as edges and bars. That is the question that we tackle experimentally in this paper. Previous thinking has centred on two main ideas: that edges and bars might correspond to (1) specific locations (peaks or zero-crossings) in the luminance profile or its spatial derivatives, after some degree of spatial smoothing, or (2) specific locations (peaks) in the local contrast energy profile of the image, after bandpass spatial filtering. Very briefly, in the first approach bars might correspond to peaks in the smoothed luminance profile and/or to selected peaks in the second derivative, while edges might correspond to peaks of gradient (first derivative) magnitude which can be also be found as zero-crossings in the second derivative. The local energy approach is motivated by the idea that image features correspond to locations where there is maximum similarity ('congruence') in local phase across a range of spatial frequencies. Such points of phase congruence give rise to peaks in a local energy measure that is computed by taking the quadratic sum (square-root of the sum-of-squares) of the outputs of even- and odd-symmetric bandpass filters that have matched spatial frequency tuning curves. The derivative-based approach in computer vision and human vision is represented in the work of Boie and Cox (1987), Canny (1986), Elder (1999), Elder and Zucker (1998), Georgeson (1994), Georgeson and Freeman (1997), Lindeberg (1998), Marr and Hildreth (1980), Watt (1988), Watt and Morgan (1985), Zhang and Bergholm (1997), while the local energy model was pioneered by Burr and Morrone (1992), Burr and Morrone (1994), Burr, Morrone, and Spinelli (1989), Morrone and Burr (1993), Morrone and Burr (1988), Morrone, Navangione, and Burr (1995), Morrone and Owens (1987), Ross, Morrone, and Burr (1989), with further theoretical developments for machine vision by Kube and Perona (1996), Kovesi (2000), van Deemter and du Buf (2000).

The two approaches have much in common, both in the general aim of producing a feature description of the image, and more specifically in the use of even- and odd-symmetric bandpass filters to do so. Derivative operators of odd order (first, third, etc.) have receptive fields with odd symmetry, while those of even order

(0, 2, 4, . . .) have even symmetry. There is some psychophysical and physiological evidence that vision might specifically employ even and odd filters (Burr et al., 1989; Field & Nachmias, 1984; Kulikowski & King-Smith, 1973; Ringach, 2002; Shapley & Tolhurst, 1973), rather than a range of filters with arbitrary or random phases (DeAngelis, Ohzawa, & Freeman, 1993; Field & Tolhurst, 1986), but this evidence does not distinguish the two approaches.

1.2. Distinguishing the two approaches

It is possible, however, to create images for which the two approaches make very different predictions about where edges and bars will be seen. Georgeson and Freeman (1997), for example, used compound gratings containing just two sinusoidal components ($f + 3f$) and had observers mark where they saw bars and edges in these 1-D images, across a range of phase relationships between the f and $3f$ components. The number and type of features seen, and the systematic way that their positions varied with relative phase of f and $3f$ components, were in good agreement with the derivative-based approach, but were poorly predicted by the peaks of local energy. This evidence clearly favoured the derivative-based approach but was not entirely convincing. Burr and Morrone (1994) suggested that the energy model is particularly successful with broad-band images containing many Fourier harmonics, because these create strong peaks in local energy (cf. Kovesi, 2000). The images used by Georgeson and Freeman were very blurred, and contained only two, low-frequency, Fourier components. These images may have been too restricted to serve as a fair test of the energy model, and so we need a more systematic study of human feature perception using broad-band images that sample widely across three important image parameters: phase, blur and contrast.

The images necessary for such a study were devised by Morrone and Burr (1988). The amplitude spectrum is that of a partially blurred square-wave grating, but with all Fourier components shifted by a constant phase angle (ϕ). Fig. 1 illustrates the family of waveforms created by different degrees of phase shift. Blur, contrast and phase can be independently varied. Note that the luminance waveform varies markedly with the phase parameter, but (by design) the energy peak location remains fixed at the point where all the component phases coincide (at $x = 0$ in Fig. 1). Formal definition of this class of images is given in Section 2.1. It is also important to note that the images have a broad-band spectrum similar to that of natural images. The amplitude spectrum of the original square-wave has the $1/f$ form that is typical of natural images (Field & Brady, 1997; Ruderman, 1997), except that in the square-wave case the spatial frequencies are discrete odd harmonics,
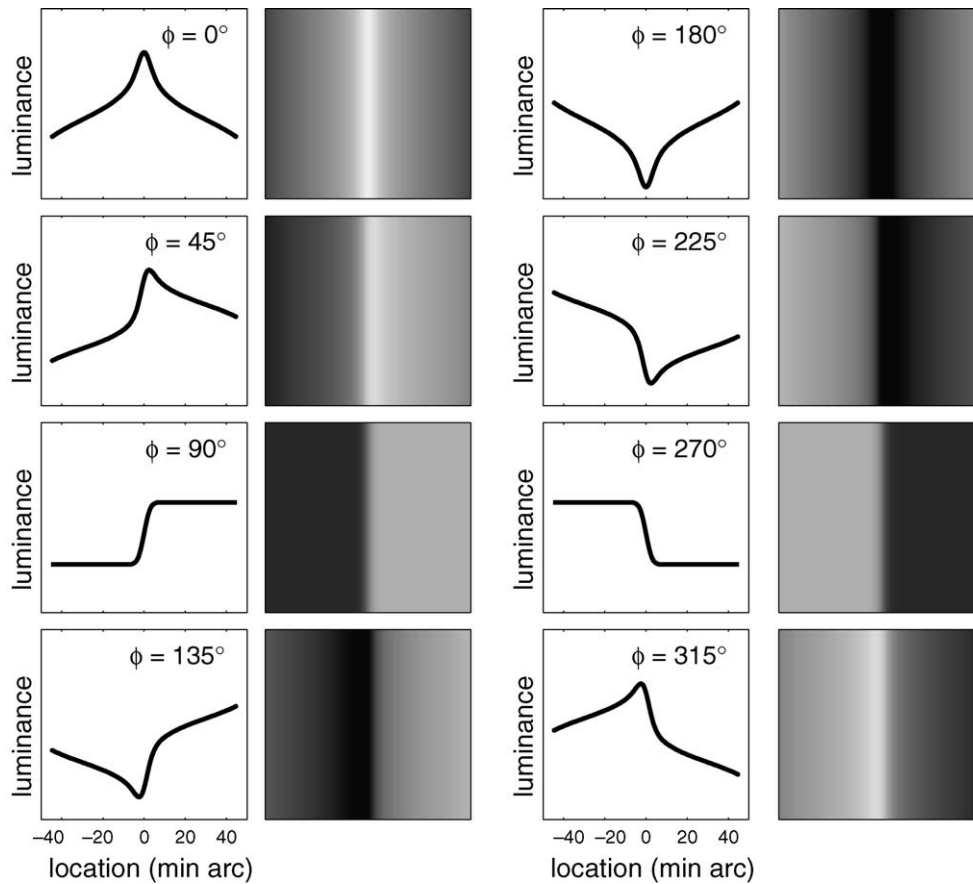
Fig. 1. Luminance profiles of the test images, for medium blur ($\sigma_b = 4$ c/deg). Each image in the set has the amplitude spectrum of a blurred square-wave, with a phase shift ϕ applied to each Fourier component (Eq. (1)). Images were 1 period (2°) wide; the central 1.5° are shown here.

$f = h \cdot f_0$, where $h = 1, 3, 5, \ldots$, and f_0 is the fundamental spatial frequency. The bandwidth of the test images also depends on the degree of blur imposed on them, but for small or moderate blurs the images can be considered spectrally broadband.

Our study thus aims to find out where people see bars and edges in this class of images, and to determine whether there is any simple stimulus-based rule that can account for their judgements. We used a feature marking method (Experiment 1) and a contour alignment task (Experiments 2 and 3) to provide converging evidence about human perception of edges and bars.

2. Experiment 1—Feature marking

2.1. Stimuli

We used 1-D vertical images equivalent to those defined by Morrone and Burr (1988). Each image had the amplitude spectrum of a blurred square wave with all Fourier components shifted by a constant phase (ϕ). The luminance at each horizontal pixel location x, is defined by:

$$L(x, \sigma_b, \phi) = L_0 + \frac{4a}{\pi} \sum_{h=1,\text{odd}}^{h<T/2} \frac{1}{h} G(hf_0, \sigma_b)$$
$$\times \cos(2\pi h f_0 (x - \mathrm{d}x) - \phi) \tag{1}$$

where T is the image width in pixels (T was fixed at 256 pixels, subtending 2°), h is the harmonic number of each Fourier component—an odd integer less than the Nyquist limit ($T/2$), L_0 is the mean luminance, a is the mean to peak amplitude of the square wave (when $\phi = \pi/2$) (Morrone & Burr, 1988), ϕ is the phase of all the cosine Fourier components at the origin and $\mathrm{d}x$ is a rigid spatial offset of the whole waveform ($\mathrm{d}x = 0$ in Experiment 1, variable in Experiments 2 and 3). The fundamental frequency f_0 was 0.5 c/deg (1 c/image). The sharpness of the stimulus is specified by σ_b, the standard deviation in c/deg of a Gaussian blurring filter defined in the frequency domain as $G(f, \sigma_b) = \exp(\frac{-f^2}{2\sigma_b^2})$. When $\phi = 0°$ (or 180°) the peaks (or troughs) of all the components coincide at the centre of the image, resulting in a triangular type of waveform (Fig. 1). When $\phi = 90°$ or 270° all the components are zero valued at the centre of the image, producing a Gaussian-blurred square-wave edge (Fig. 1).

A set of eight images was used for each experimental session. All images in the set had the same sharpness (σ_b) but different phase shifts ($\phi = 0°$, $45°$, $90°$, $135°$, $180°$, $225°$, $270°$ and $315°$). Examples of the 1-D luminance profile and corresponding grey scale image for each phase shift when $\sigma_b = 4\,c/deg$ can be seen in Fig. 1. Note that for a given sharpness there are only three basic wave shapes in the set of images, at phases $0°$, $45°$ and $90°$; waveforms at the other five phases are reflections and/or inversions of the first three. As described previously, a is the mean-to-peak amplitude of the square wave. It is also the standard deviation of the luminance profile of the waveform. Because of Parseval's theorem this value does not change with phase and it can be used to derive the root-mean-squared contrast (rms contrast) as a/L_0 which is constant within a particular image set.

2.2. Apparatus

The experimental stimuli were generated in NIH Image software [http://rsb.info.nih.gov/nih-image/] using Eq. (1). The experiment was run on a Macintosh computer with a calibrated, gamma-corrected Apple 15″ CRT display. Control of image contrast and linearisation of the relationship between the digital signal and screen luminance were achieved by manipulating the lookup tables (LUTs). The mean luminance of the screen was $60\,cd/m^2$. The experimental display was viewed binocularly in a dimly lit room at a viewing distance of 2.92 m. The head was restrained by a chin and forehead rest. The width and height of each test image were 256 pixels (10.2 cm) which subtended a visual angle of $2°$. One pixel subtended $0.47'$. Each image was set in a uniform grey (mean luminance) background ($5° \times 3.73°$).

2.3. Procedure

Six subjects took part in the experiment using binocular viewing and normal spectacle correction where required. There were four naive observers and two experienced observers (GSB, MAG). For each subject the experiment consisted of 15 sessions—one for each level of stimulus sharpness ($\sigma_b = 1, 2, 4, 8, 16\,c/deg$) and rms contrast (24%, 12%, 6%). During a session each of the eight phases was presented five times, in random order.

Each image was flashed repeatedly for 300 ms with a blank (mean luminance) inter-stimulus interval of 700 ms. The purpose of the intermittent presentation was to prevent the build-up of afterimages that can cause an apparent instability ('monocular rivalry') of the stimuli at certain phases ($45°$, $135°$, $225°$ and $315°$) (Georgeson & Freeman, 1997). A thin red vertical marker (1×24 pixels, $0.47 \times 11.28'$) was superimposed onto the image and the subject could move the marker horizontally across the image using the mouse. When the

marker appeared to be centred over the next bar or edge the subject would click the mouse button to report the feature location and then select from a set of four icons to indicate the type (bar or edge) and the polarity of the feature seen. After selection the marker remained at the same location. The subject would continue moving the marker across the image in the same direction marking other features in the same fashion. Subjects were free to mark as many (or as few) features as they wished. When the marker reached the end of the selection area a new image from the set was presented. The direction of marker movement (left to right or right to left) was alternated between test images. For all images the selection area was limited to ±0.5° from the image centre. This restriction did not exclude any parts of the image from the experiment, because the parts excluded from one trial would be those included in another trial at a different phase. Subjects could not make selections outside this area and could not select the same feature twice by making selections in the 'wrong' direction. Free eye movements were allowed.

3. Experiment 1—Results

3.1. Consistency of reported locations

The left column in Fig. 2 shows the features marked by a single subject (NTB) across five different trials for phases of $0°$ and $45°$ at the medium stimulus sharpness ($\sigma_b = 4\,c/deg$) and medium rms contrast (12%). The following symbols are used to indicate the feature type and polarity: filled squares are dark bars (D), open squares are light bars (L), filled triangles are dark to light edges (DL) and open triangles are light to dark edges (LD). It is clear that for these images the subject consistently reported the same feature type, polarity and spatial relationships of features across the five repetitions. A single central edge was reported at phases $\phi = 90°$ and $\phi = 270°$ and a central bar flanked by two edges was reported for all other phases.

This pattern of results was consistent across subjects, and across different levels of stimulus sharpness and contrast. In particular there was little or no effect of contrast level. The only exceptions were for the most blurred images ($\sigma_b = 1\,c/deg$) at the lowest contrast (6%). Here subjects DJC and MKB reported a single edge when $\phi = 45°$, $135°$, $225°$ and $315°$, GSB reported a single bar when $\phi = 315°$ and MAG reported a bar and edge when $\phi = 45°$ and $315°$.

The right column in Fig. 2 illustrates the average positions reported by individual subjects for phases of $0°$ and $45°$ when $\sigma_b = 4\,c/deg$ and rms contrast = 12%, along with the group mean across subjects. The standard deviation of reported locations across the five repetitions, averaged across subjects, stimulus sharpness and
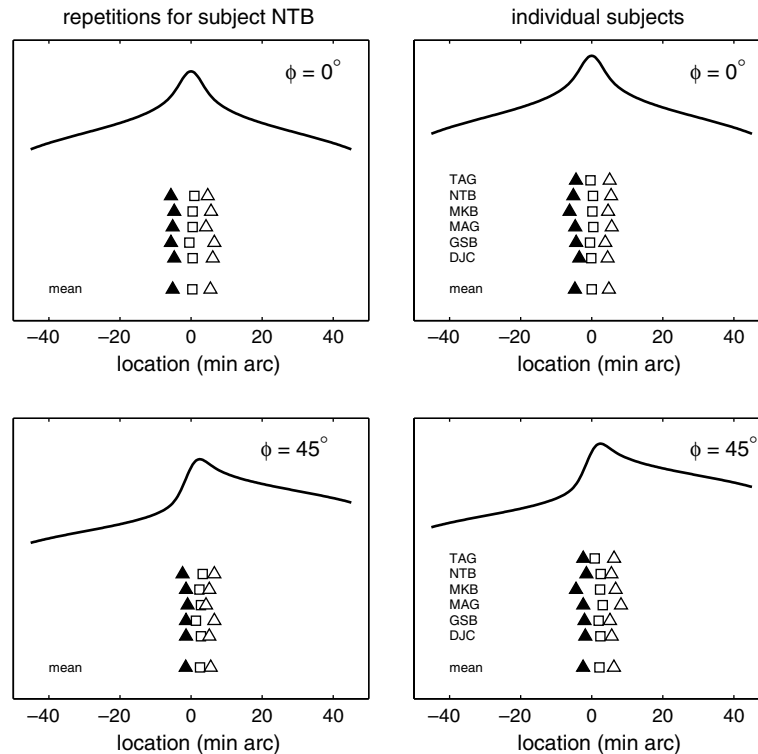
Fig. 2. Locations of perceived features marked for phases 0° and 45° when σ_b = 4 c/deg and rms contrast = 12%. Filled triangles: dark-to-light edges; open triangles: light-to-dark edges; open squares: light bars. Left column: features marked by one subject (NTB) in individual trials; bottom row in each panel shows the mean feature location for NTB over five trials. Right column: means for individual subjects; bottom row in each panel is the group mean. Consistency within and between individuals was very high.

contrast was 0.90′, and the average correlation of the reported locations between all pairs of subjects was 0.98. Because of this robust similarity of observations across subjects all further analysis was carried out on the average feature locations taken across all six subjects. At the lowest contrast (6%) and sharpness (1 c/deg) there was some inconsistency between subjects (described above), so a feature was included in the average only when at least four out of six subjects reported it.

3.2. Pattern of perceived locations across phases

Fig. 3 shows a plot of the average marked feature locations against the spatial phase of each image for each level of sharpness (σ_b) and contrast. Because there was no systematic effect of contrast level, the symbols for lower contrasts are mainly obscured behind those for the highest contrast. The location of each feature shifts with phase (ϕ), tracing out a path which we call a *feature trajectory* (Georgeson & Freeman, 1997). The bar trajectories are centred at the image origin ($x = 0$) for the triangle-wave phase condition (0°, 180°) and move away from the centre as the phase shifts towards 90° or 270° (−90°), but the bars disappear altogether at ±90° where only a single edge is seen. The edge trajectories are centred at the image origin for the blurred square-wave condition (±90°) and move away from cen-

tre as phase shifts toward 0° and ±180°. Note that, when the phase is intermediate (±45°, ±135°), the reported locations of both bars and edges are displaced systematically from the origin. As stimulus sharpness increases, the trajectories become shallower, implying that the shift in reported location with phase is smaller for sharper stimuli.

3.3. A simple rule for feature locations

We explored the data in a variety of ways (discussed in more detail later) in search of a stimulus-based definition for human feature perception. One simple rule can account for essentially all of these data, and is easily stated as:

Rule 1. Human observers see light and dark bars at peaks and troughs in a slightly smoothed version of the luminance profile, and they see edges at points of steepest gradient in that smoothed profile.

The pattern of feature trajectories predicted by Rule 1 is shown as solid lines in Fig. 3. This rule predicted the marked locations very well at all phases, and all levels of blur and contrast (except in those few cases of large blur and low contrast already noted). The space constant (σ_x) of the Gaussian blur kernel used to smooth the
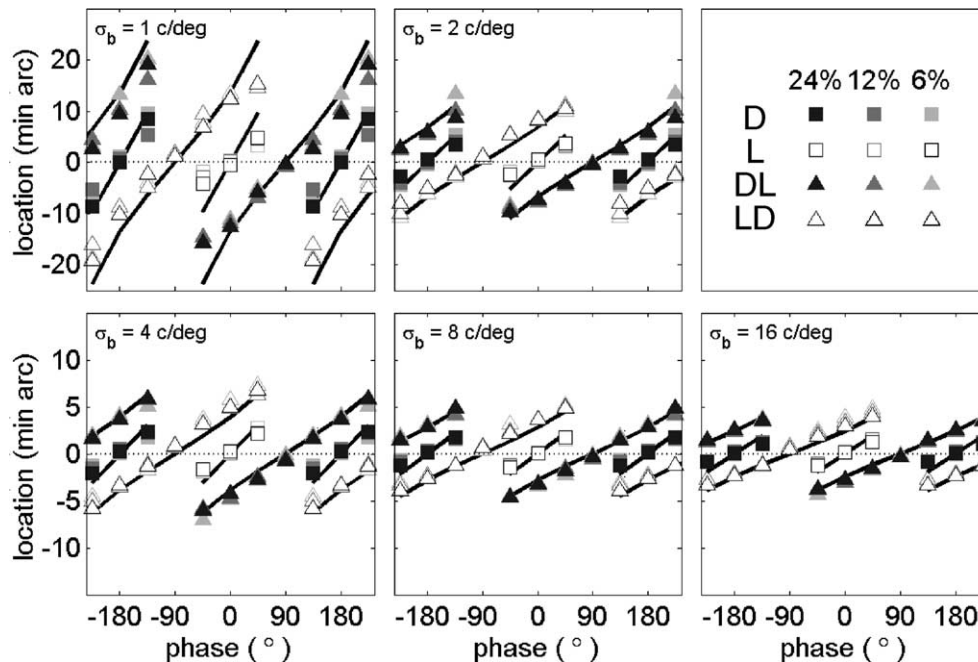
Fig. 3. Feature trajectories for each level of sharpness (σ_b) and contrast (24%, 12%, 6% rms). Group mean feature locations are plotted against component phase. Filled squares: dark bars; open squares: light bars; filled triangles: dark-to-light edges; open triangles: light-to-dark edges. Symbol contrast represents image contrast (see top right); low contrast symbols are mainly obscured by higher ones, reflecting little or no effect of image contrast on feature locations. Solid curves: features predicted by a simple rule (Rule 1—see text). Dotted horizontal line is the energy peak location for all phases. Phase is a circular dimension, and so data at 180° should be envisaged as wrapping around to join those at −180°. Note change of vertical scaling in lower row.

luminance profile was small, 1.7′, corresponding to a low-pass Gaussian filter in the frequency domain with $\sigma_f = 5.5$ c/deg. The overall goodness-of-fit is excellent. It is clear that the rule predicts the observed features well for all levels of stimulus sharpness. It predicted the correct number, type and polarity of the features and placed each feature close to the corresponding observed feature. The only exceptions were when the stimulus was most blurred ($\sigma_b = 1$ c/deg), rms contrast was low (6%) and $\phi = 45°$ or 315°. In these two conditions the rule predicted two edges (as it did for all other levels of blur and contrast) but most subjects reported only one edge. Our best guess is that the unseen edge fell below a local contrast threshold in this low-contrast condition.

3.4. Evaluation of the energy model

The images in this study were designed specifically (by Morrone & Burr, 1988) to produce a single energy peak at the centre of the display screen. For each image, the Fourier components of the image have a common phase when referenced to that point, and so the prediction of the energy model is a simple one: every image should have a single salient feature at the same central point. The feature trajectory should be a single horizontal line, but it is not. There are three, almost parallel, non-horizontal, feature trajectories representing a bar

flanked by two edges (Fig. 3). The slopes of these trajectories increase as blur increases, but they are invariant with contrast. For this image set, the energy model does correctly predict the edge at 90° and 270°, and the bar at 0° and 180° phases. But it predicts no phase-dependent variation in feature locations, and so it captures none of the systematic variation observed in the data as phase varied. Likewise, since it predicts only a single feature in each image, it does not describe the consistent finding, across all blurs, that every visible bar is flanked by two edges of opposite polarity.

The relation between energy peaks and observed features is illustrated in Fig. 4. The solid curve shows a portion of the luminance profile of a medium-sharp image (4 c/deg) when $\phi = 45°$, along with the location of the central bar and two flanking edges marked in the experiment (rms contrast = 12%). The luminance profile is skewed, and the observed light bar is shifted away from the central location ($x = 0$), but sits near the luminance peak. The energy profile (Fig. 4(A)) predicts that there should be a single feature located at the centre of the image but three features were consistently observed, none of which was located at the centre. The arrow-plots in Fig. 4(C) show the local phase (arrow orientation) and amplitude (arrow length) of the Fourier components at the observed feature locations and at the energy peak. At the marked feature locations the component phases vary greatly with frequency, while at the point of phase
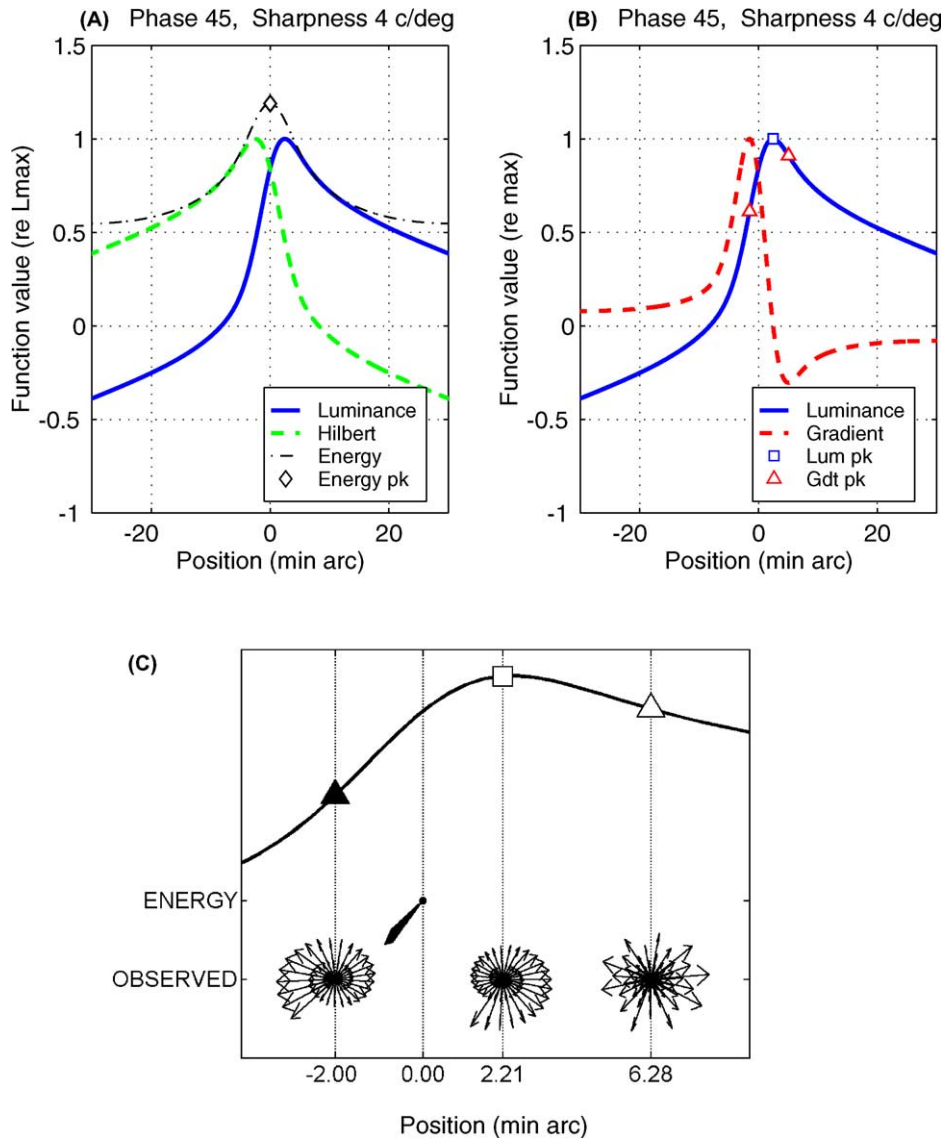
Fig. 4. Energy and phase congruence. (A) Luminance profile, Hilbert transform and energy profile of the test image for $\phi = 45°$, $\sigma_b = 4$ c/deg. (B) Gradient profile (dashed curve) for the same luminance waveform (solid curve). Symbols mark the luminance peak and the points of steepest gradient. (C) The group mean edge positions (triangles) and bar position (square) marked in the experiment (at 12% rms contrast). Arrow-plots show phase (arrow orientation) and amplitude (arrow length) of Fourier components at the observed feature positions and at the energy peak position, $x = 0$. Congruence of component phase did not predict feature locations, and perceived features did not have phase congruence.

congruence no feature is marked. This example is representative of the whole dataset for these critical intermediate phases (±45, 180 ± 45). Figs. 3 and 4 imply that the congruence of phases at a particular location is neither necessary nor sufficient for the perception of features. It appears that the inability of the energy model to predict these marked features is not restricted to blurred or narrow-band images (see Section 1), but is more general.

4. Experiment 2—Contour alignment

Burr and Morrone (1994, p. 139) argued that the perceived structure of images is determined to a large extent by the visually salient features formed where the harmonics come into phase with each other. In Experiment 1 observers had to locate bars and edges, but they did not have to judge shape or spatial relationships. An interesting possibility, therefore, is that when asked to identify local features explicitly, observers mark luminance peaks and gradient maxima (they follow Rule 1), but if we had asked them to make some more global judgement about shape or form we would have found that energy peaks were key features for such perception of structure. An important aspect of perceived structure, much studied in recent years, is the perception of collinearity—the alignment (or misalignment) of contours extended across

space—and so perhaps the alignment of energy peaks would determine this kind of perceived structure. This was tested in Experiment 2, using a 3-element alignment task with the same family of spatial waveforms as Experiment 1.

The logic was straightforward, and is illustrated in Fig. 5(A) and (B). A pair of outer contour segments had either 0° or 180° phase (light or dark bars), while the centre element was phase shifted by ±45° from the outer elements. If perceived alignment is determined by energy peaks, then the three elements should appear aligned, because the phase shift of the centre segment does not displace the energy peak. However, if luminance peaks (and/or gradient peaks) determine perceived alignment, then the phase shift in the centre will induce misalignment. This can be quantified by finding the spatial offset dx of the centre segment that is required to restore perceived alignment.

4.1. Stimuli and procedure

Images were generated in *Matlab* software on a Macintosh G4 computer, and displayed on an Iiyama Pro454 19″ monitor at a framerate of 85 Hz, using the Pelli–Brainard *PsychToolbox* software (Brainard, 1997) [http://psychtoolbox.org/]. The display was carefully gamma-corrected, and had a mean luminance of 77 cd/m². Where possible, the stimulus conditions were the same as in Experiment 1: luminance profiles were defined by Eq. (1), the image width was 2°, fundamental frequency was 0.5 c/deg. Only the highest rms contrast (24%) was tested. The remainder of the screen surrounding the test image subtended 8.9° × 6.5°, and was held at mean luminance throughout the session. In different blocks of trials the sharpness parameter (σ_b) was 2, 4 or 8 c/deg and the outer elements (Fig. 5(A) and (B)) had phases of 0° or 180°, with dx = 0. The uniform grey (mean luminance) strips between the three contour seg-
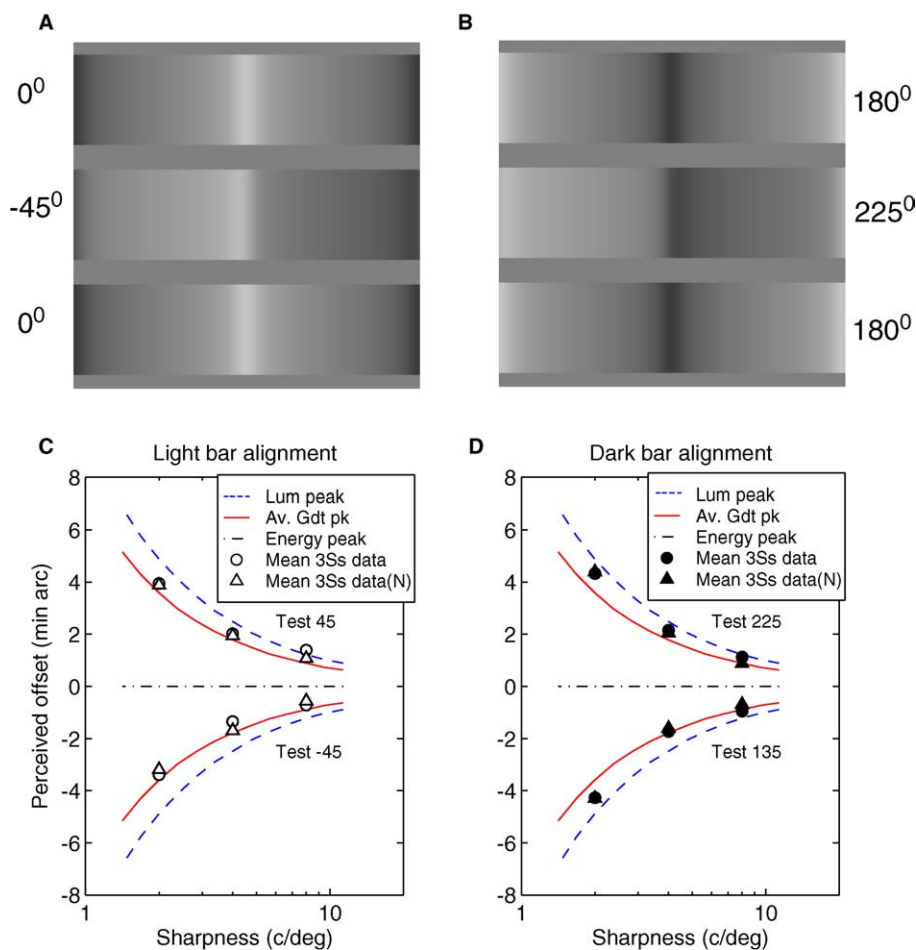

Fig. 5. Experiment 2. (A, B) Arrangement of images used in the contour alignment task. Phases of the outer contours were 0 (A; light bars) or 180 (B; dark bars). Phase of the test contour (centre) was shifted ±45° from the outer ones. Spatial displacement (dx) of the centre varied from trial to trial. These images (for σ_b = 4 c/deg) show dx = 0, where the energy peaks are aligned, but the reader may observe that the centre appears offset to the left (A) and right (B). (C, D) Mean values of perceived offset from three observers for the contour alignment task with grey dividing strips (circles; Experiment 2) or binary noise strips (triangles; Experiment 3).

ments were 8.4′ high. The centre segment had its phase shifted by +45° or −45° relative to the outer segments. This meant that the polarity and perceived features of the centre and outer elements were similar (see Fig. 5(A) and (B)) but the energy peak and luminance peak could be dissociated. The phase offset (±45°) of the centre element switched at random from trial to trial, as did its spatial offset dx. Values of dx were drawn at random from sets of seven that were chosen for each level of blur and direction of phase offset, on the basis of pilot data.

Within a trial, images were presented once for 0.5 s. The observer was asked to fixate centrally (without a fixation point) and had to indicate with a key press whether "a salient feature roughly in the middle was to the right or left of the outer pair". The nature of the feature was left undefined, and no feedback was given. Twenty trials were run for each offset dx, and the six blocks of 280 trials were run in random order. Thus 1680 trials were run for each of three observers in single sessions lasting about 1 h. Viewing was binocular from
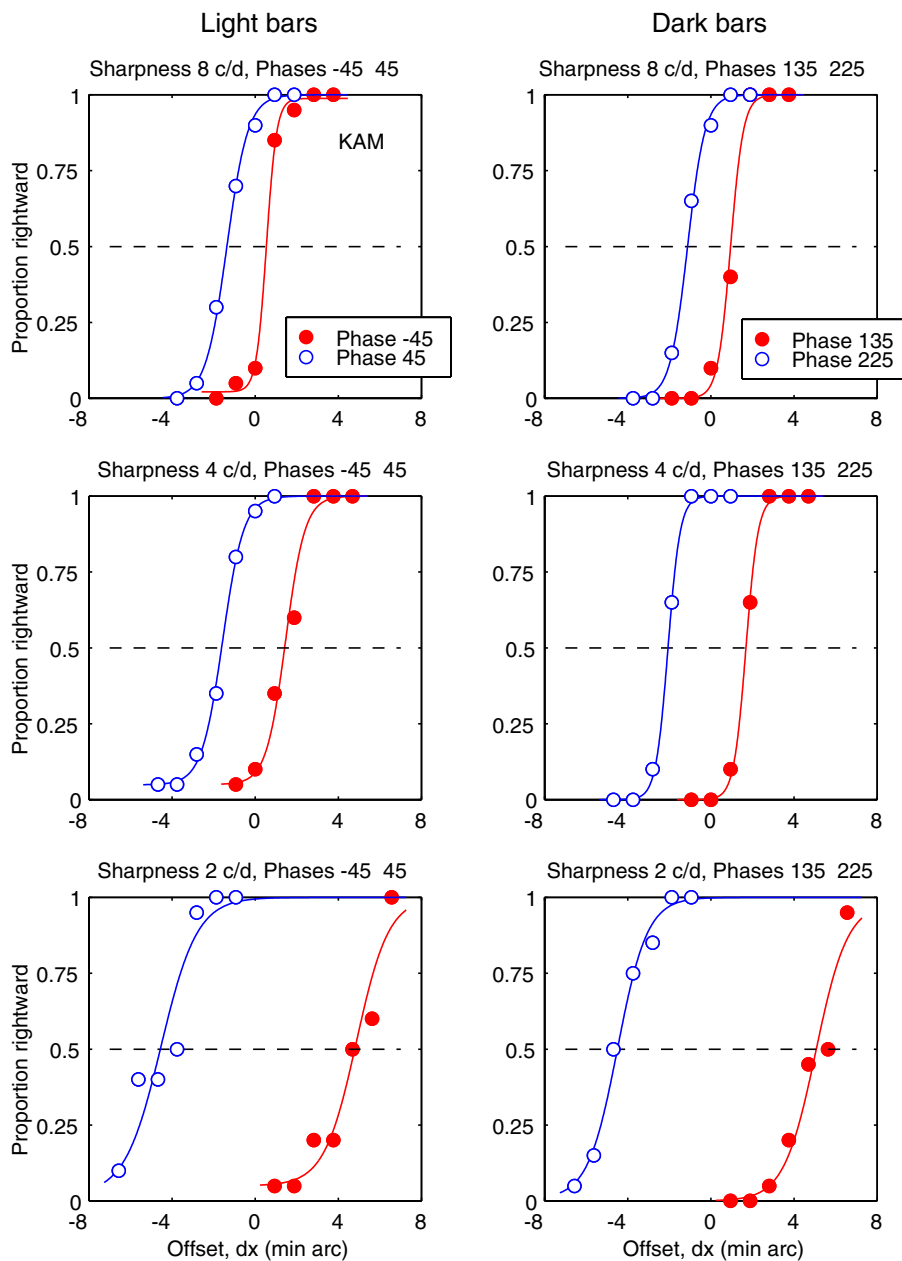

Fig. 6. Experiment 2. Data and fitted psychometric functions for one observer (KAM), illustrating the large difference in point of subjective alignment [where p(rightward response) = 0.5] when the test (centre) phase is shifted by +45° (open circles) versus −45° (filled circles), relative to the flanking contours. Sharpness (2, 4, 8 c/deg) increases from bottom to top row. Left column: light bars (flank phase 0°). Right column: dark bars (flank phase 180°).

a distance of 229 cm. Subjects were author MAG, an experienced but uninformed observer (KAM) and a naïve observer (CXP). Prior practice varied from several hours (MAG), through 1 h (KAM) to about 5 min (CXP).

4.2. Experiment 2—Results

Data were plotted as the proportion of rightward responses for each blur and centre phase as a function of the centre offset dx. Maximum-likelihood psychometric functions (logistics) were fitted to each subset of data, using the *Psignifit* software (Wichmann & Hill, 2001) [www.bootstrap-software.com/psignifit] as shown for KAM in Fig. 6. The pattern of data was very similar for the other two subjects, but with shallower psychometric functions for the naïve subject. Fig. 6 (open symbols) shows that when the centre element phases were shifted to the right (centre phase 45°, or 225°), the whole waveform had to be shifted to the left (dx < 0) to achieve subjective alignment with the outer elements, and vice-versa (filled symbols). This implies that rightward (or leftward) phase shift led to a perceived positional shift to the right (or left) respectively. The reader may be able to see this in the illustrations of Fig. 5(A) and (B). The degree of positional shift increased with the level of blur.

The steepness of the psychometric functions illustrates the finding that, on average, the positional shift was 3–5 times greater than the just-noticeable difference (JND) in centre element position (see Table 1). JNDs also increased with blur (Table 1), and JNDs for individual Ss were associated with the observer's level of practice (MAG: 0.25; KAM: 0.52; CXP: 1.1′, geometric means over blur and phase). The precision of these positional judgements confirms our subjective impressions that the task was natural and unambiguous.

To summarize the results, the positional shifts (50% points on the fitted psychometric functions) were averaged across subjects, and plotted against the image sharpness (Fig. 5(C) and (D), circles). Dashed curves in Fig. 5(C) and (D) show that for both the light bars (C) and the dark bars (D) the perceived offsets fell fairly close to those predicted by the positions of luminance

peaks (or troughs) in the displayed image. However, the solid curves reveal that the perceived offsets were even closer to the positions defined by the midpoint between pairs of gradient peaks that flanked the luminance peak (cf. Fig. 4(B)). Taking the findings of Experiments 1 and 2 together, we conclude that each bar has two edges located very close to gradient peaks, and that perceived alignment is closely associated with the average alignment of these edges.

5. Experiment 3—A control experiment

Up to this point, we followed Morrone and Burr (1988) in calculating 1-D energy profiles from horizontal cross-sections of the luminance profile. This seems appropriate for Experiment 1, where the images were genuinely 1-D, with no modulation on the vertical axis. However, it may not be appropriate for the images of Experiment 2 which had 2-D variations in luminance. Specifically, the grey strips that separate the three vertical contour segments created horizontal edges, and the local contrast of these edges depends on the adjacent luminance profile. When phase (and hence luminance peak) shifts to the left (centre of Fig. 7(C)) so the point of maximum contrast on the abutting horizontal edges also shifts in the same direction. This produces a cue in the output of horizontally oriented energy mechanisms, as illustrated in Fig. 7(D). Bright peaks of energy at the ends of the central test contour are shifted to the left, relative to those created by the outer contours. This was evident at all spatial scales for horizontal mechanisms, and so for brevity we summarized this pattern in Fig. 7 by the sum across all scales. Note that in Fig. 7(D) the (fainter) vertical ridges of energy, arising from vertical mechanisms responding to the body of each test contour, remain perfectly aligned—unaffected by phase shifts, just as in the 1-D analysis.

In short, then, a 2-D energy model might account for the perceived offsets of Experiment 2 if observers based their alignment judgements on those cues that arise from horizontal mechanisms. We tested this in a replication of Experiment 2, but with the addition of binary noise to the grey separating strips (Fig. 7(E)). At all spatial scales, noise of this size and contrast scrambled the offset cue described above (see Fig. 7(F)). We confirmed by computation with many different noise samples that this scrambling was effective at all scales even when four oriented energy mechanisms were considered (0°, 45°, 90°, 135°). So, for an energy-based observer, this noise should eliminate the perceived offset associated with phase shifts. The only alignment cue is now the ridge of energy from vertical mechanisms, and this, as we have seen, is phase-invariant and predicts no perceived offset.

Table 1
Experiment 2. Group mean positional shift for 45° phase shift (PSE) and position discrimination (JND)

Sharpness σ_b, c/deg	Mean shift in PSE (min arc)		GeoMean JND (min arc)		Ratio PSE/ JND	
8	1.04	*0.80*	0.39	*0.58*	2.71	*1.38*
4	1.81	*1.83*	0.47	*0.68*	3.83	*2.70*
2	3.99	*3.94*	0.79	*1.17*	5.05	*3.37*

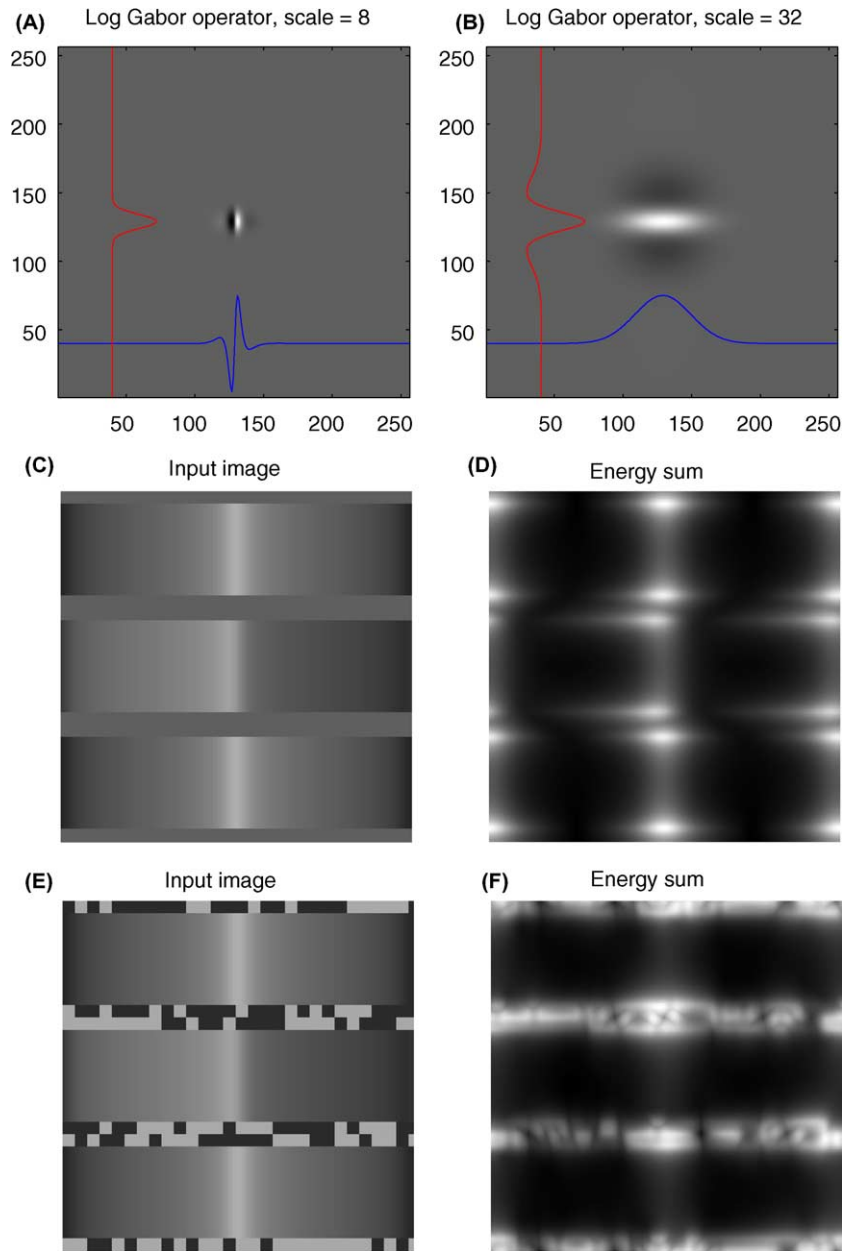Corresponding values for Experiment 3 are in *italics*.

Fig. 7. 2-D energy maps. (A, B) Examples of the filter kernels used to compute energy maps at different scales and orientations. Filters had the log Gabor form defined in the Fourier domain by Burr and Morrone (1992). At each scale and orientation, a pair of odd and even filters was used to compute energy maps (the quadratic sum of the pair of filtered output images) in the usual way. Peak sensitivity and octave bandwidth were the same for all filters. (C) Test image like Fig. 5(A), where the centre contour has $-45°$ phase shift. (D) Energy summarized as the sum over three scales (8, 16, 32; centre frequencies 16, 8, 4c/image) and two orientations (0°, 90°). Note the offset in energy peaks where the test contour meets the grey dividing strip. This offset cue arises in the horizontal filters at all scales (see text). (E) As (C), but with binary noise added to the grey dividing strips. (F) Note how the offset cue in the energy map is completely scrambled by the addition of this noise.

5.1. Display and procedure

Display parameters and experimental procedures were the same as Experiment 2, except that (i) the display screen was an Eizo 6500 greyscale monitor with a mean luminance of $51 \, cd/m^2$, viewed from 259 cm; (ii) binary noise, with square elements 4.2′ wide and contrast 0.5, was added to the grey separating strips (Fig. 7(E)), with new noise samples on every trial; (iii) subjects were author MAG and KAM as before, with a different naïve observer (RJS).

5.2. Results

Results were strikingly similar to those of Experiment 2, in both the individual psychometric functions (not

shown) and the mean perceived offsets (Fig. 5(C) and (D), triangles). Clearly, the noise had no effect on the observers' mean judgements of alignment, and so we can be confident that peaks of contrast along the horizontal edges were not the cue used to judge alignment in this task.

6. Discussion

6.1. A simple rule for edge and bar locations

For the class of images studied here, we found that a very simple rule (Rule 1) provides a good account of marked feature locations, as well as feature type and polarity, across all phases and for all degrees of blur. After a small amount of spatial smoothing, bars were located at peaks and troughs in the (smoothed) luminance profile while edges were located at points where the luminance was changing most steeply across space (peaks in gradient magnitude). The latter has been the underlying assumption of many edge-finding techniques in machine vision (Canny, 1986; Marr, 1982; Marr & Hildreth, 1980). Our results and analysis provide strong evidence that this assumption also holds for the human visual system.

In Marr's (1982) terms, these findings tell us something about human vision at the level of computational theory (what is to be computed), rather than how it is computed. For example, the luminance peaks and troughs could be found directly by searching over an internal representation of the luminance profile, but equivalently they could be found as zero-crossings in the output of a gradient (first-derivative) operator. Similarly, edges could be found as peaks in the output of the gradient operator (Canny, 1986) or by finding zero-crossings in the second-derivative output (Marr & Hildreth, 1980). Many models have implemented such ideas using derivative filters at multiple scales (e.g. Watt & Morgan, 1985) or with local, adaptive selection of the filter scale (e.g. Elder & Zucker, 1998; Lindeberg, 1998). Our results do not bear directly on the merits of these various implementations (they are not 'analyzer-revealing') (Graham, 1989), but they do impose constraints on what any model intended to account for human vision must produce as output. In particular the results suggest that peaks of local contrast energy are not what the human observer uses to locate luminance features. At the critical intermediate phases (±45, 180 ± 45), energy peaks do not predict the correct number or position of features. This conclusion echoes and strengthens our earlier one (Georgeson & Freeman, 1997) because the stimulus set in the present study was designed to be a fairer, more complete test of the energy approach. Of course, this conclusion is limited to the class of smooth, noise-free, 1-D images used in these

experiments, and we re-emphasize that Rule 1 does no more than describe succinctly the features that are reported for this class of images.

6.2. Is there a better rule?

It is reasonable to ask whether other rules could work as well as Rule 1, or perhaps better. To explore this systematically, we took Rule 1 to be a special case of a more general scheme for potential feature descriptions (cf. Georgeson & Freeman, 1997). We consider first bars, then edges. In general, bars might correspond to peaks and troughs in the output of an even-symmetric filter applied to the luminance profile, with the filter's amplitude spectrum F defined by:

$$F(f) = f^p \exp\left(\frac{-f^2}{2\sigma_f^2}\right) \quad (2)$$

In terms of Eq. (2), Rule 1 asserts that $p = 0$ (i.e. simple smoothing by a Gaussian filter) but when $p > 0$, this function combines high-pass filtering and Gaussian smoothing. Second-derivative (Laplacian of Gaussian) filtering (Marr & Hildreth, 1980) is given by $p = 2$. Extensive numerical computations showed that $p = 0$ gave the best fit to the feature-marking data for two reasons: (i) when $p > 0$ there were too many peaks and troughs in the filter output, and (ii) when $p < 0$ the number of peaks and troughs matched the observed bars but the positions were less accurate than when $p = 0$. Fig. 8(B) shows that when p was set to 0, and σ_f varied, there was usually no optimum filter scale for predicting bar locations, but the rms positional error was large when too much smoothing was applied, and approached a lower bound when σ_f was greater than 5–8 c/deg. Thus slight, or no, smoothing gave similarly good predictions for bar locations, provided $p = 0$. That is the first part of Rule 1.

To conduct a similar analysis for edges, we applied Eq. (2) as an odd-symmetric filter instead of an even one, and searched for peaks in the magnitude of the filter output (i.e. rejecting peaks and troughs that are minima in magnitude (Clark, 1989)). Rule 1 now asserts that $p = 1$ (the Gaussian derivative operator). The outcome of this analysis was very like that for bars. When $p > 1$ too many edges were produced, and when $p < 1$ the correct number of edges was produced but the positional error was greater than for $p = 1$. Hence within this simple peak-finding scheme the Gaussian derivative operator was uniquely the best filter for predicting the observed edge locations. Fig. 8(A) shows that in this case the data for the sharper images ($\sigma_b = 4$, 8, 16 c/deg) yielded a minimum of positional error when the smoothing filter had $\sigma_f = 5.5$ c/deg, corresponding to a Gaussian blur space constant $\sigma_x = 1.7'$. For this reason we chose $\sigma_x = 1.7'$ as a single figure-of-merit for Rule
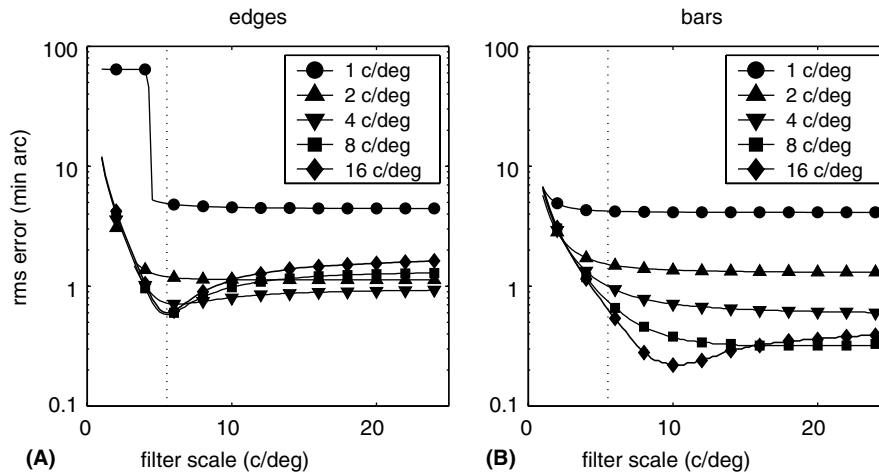
Fig. 8. Positional accuracy of Rule 1 (goodness of fit expressed as rms error, min arc) for different degrees of spatial smoothing, for (A) edges and (B) bars. Image contrast was 12% rms; image sharpness (σ_b) is shown in the key. Sudden rise in error measure for edges (1 c/deg) occurs because a fixed high penalty value was added to the error measure when an observed feature was not predicted. Vertical dotted line marks the degree of smoothing used in Fig. 3.

1, but any value from 0 to 2′ would serve almost equally well for the dataset as a whole.

Compressive non-linearity in the response to luminance can shift the perceived location of high contrast, blurred edges towards the darker side (Georgeson & Freeman, 1997; Mather & Morgan, 1986). There is a hint of that in the present data since, in Fig. 3, light-to-dark edges (open triangles) tended to lie a little above the fitted line, implying a rightward shift, with the reverse for dark-to-light edges. However, the effect was small and incorporation of a luminance non-linearity did not improve the overall fit of Rule 1 to the data, and so for simplicity we do not consider it further.

In summary, this more general analysis adds considerably to the value of Rule 1. We assumed that features are found locally as peaks in the output of some spatial filter. In a broad search of even and odd filter parameters, we did not find any better rule to describe the perceived features for this set of images. We had anticipated that there might be a different optimum spatial scale of smoothing for each level of image blur, but this was not so. Even for the most blurred images, a small amount of smoothing, or none at all, was still best. This in no way excludes multi-scale models of spatial vision, because Rule 1 is not intended as a model of visual mechanisms. But it does imply that any adequate model should produce behaviour consistent with Rule 1.

What is a bar? The analysis so far suggests that perceived bars are close to luminance peaks, but if the local-peak-finding assumption is dropped, we can see that there is another way in which bars may be found. Inspection of Fig. 3 shows that each observed bar lay almost exactly half-way between the pair of observed edges flanking the bar. In fact, the midpoint between observed edges was an almost perfect predictor of judged

bar location, as the linear relation between observed bar locations (y) and the mid-point of observed edge locations (x) was $y = 1.0458x + 0.0035$, with a correlation of 0.995 ($n = 88$; data from Fig. 3). Thus, while the observed bars could be based on luminance peaks, edge computation (followed by some form of bisection judgement) is also sufficient to account accurately for the bar-marking data as well as the alignment data (discussed later in Section 6.5). On this view, a bar is the region between two edges of opposite polarity. Neither of these rules fully explains bars, however, because neither predicts Mach Bands, discussed next.

6.3. Mach Bands

Mach Bands are bars perceived where there is a fairly sharp transition between two different luminance gradients. An increase in gradient yields a dark bar, while a decrease gives a light bar (Ratliff, 1965). In general, Mach Bands do not correspond to peaks in the luminance or gradient profiles. Mach Bands do, however, coincide with energy peaks measured at relatively small spatial scales, and the measured contrast sensitivity for seeing the bands is well predicted by the energy model (Ross et al., 1989). That notable success must be weighed up against the present finding that energy peaks are, in general, inadequate predictors of feature identity and location.

Mach Bands also coincide with peaks in the second derivative, as Mach himself observed, but here any adequate model must also account for the *absence* of Mach Bands in various images where second-derivative peaks are prominent. It has often been noted that Mach Bands are not seen on sharp step edges (Ratliff, 1984; Ross, Holt, & Johnstone, 1981; Thomas, 1965). Our data

(Fig. 3) show that this absence is more general than previously realized. At phases of ±90°, the luminance profile is a Gaussian-blurred step edge, with prominent peaks in the second-derivative, but observers never marked bars on these images for any degree of blur. Only a single isolated edge was reported. Thus Mach Bands are absent not only from sharp edges but from *all* Gaussian-blurred edges. The energy model readily predicts this absence, but it poses a strong challenge to simple derivative-based models for the perception of bars. We note, however, that the parsing rules in the MIRAGE model, discussed next, were structured to accommodate this finding.

6.4. MIRAGE

MIRAGE (Watt, 1988; Watt & Morgan, 1985) is a model of the mechanisms that lead to a feature description, and is based on second-derivative filtering at multiple spatial scales. Although model implementations were not our main concern here, it was of interest to compare the performance of Rule 1 with an elaborated model such as MIRAGE. First the signal was filtered by a set of Laplacian-of-Gaussian ($\nabla^2 G$) filters of different scales. Specifically we used five channels at octave intervals, the space constant of the smallest being 0.35′ (Watt, 1988), equivalent to a filter scale (σ_f) of 27 c/deg. The outputs from the different channels were then combined so that all the positive responses were added together to give a single positive stream and all the negative responses were added together to give a single negative stream. The separation of the negative and positive portions of the filter outputs is an important feature of the MIRAGE model. The pattern of zero-bounded re-

gions and non-active regions across the positive and negative streams was analysed using MIRAGE's parsing rules to locate and identify features and determine their polarity and relative strengths. An edge was located at the centroid-of-centroids between adjacent negative and positive zero-bounded regions that have no activity on the other sides. A bar can be located either at the centre of mass of an isolated zero-bounded region or at the centre of mass of a zero-bounded region flanked by two zero-bounded regions of the opposite polarity.

Fig. 9(A) compares the feature trajectories predicted by MIRAGE and by Rule 1 for stimulus sharpness of 4 c/deg and rms contrast 12%. Both predict the correct number, type and polarity of features. Both perform well for bars, but for edges the location errors of MIRAGE are worse as ϕ moves away from the 'pure' edge conditions (90° and −90°), while the predictions of Rule 1 remained close to the observed locations.

A defining feature of the MIRAGE implementation is that the outputs from filters at multiple scales are combined after half-wave rectification, but *before* any local primitives are extracted. Fig. 9(B) shows the RMS error for edge locations generated by Rule 1 and by two versions of MIRAGE—one using five channels as described above, the other using a single channel with the same filter scale as Rule 1. There are two important points to note. Firstly, Rule 1 has about three times better positional accuracy over all levels of stimulus sharpness. Secondly, there is little difference in performance between the MIRAGE implementation using a single filter and the version using multiple filters. This suggests that the combination of different channels prior to primitive extraction has little influence on the model's localisation of features in these stimuli, even though image
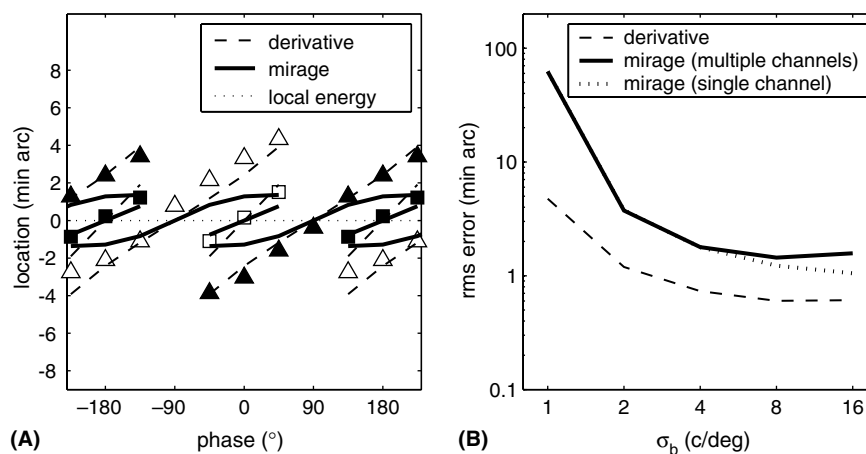

Fig. 9. Comparison with MIRAGE. (A) Observed locations of edges and bars (symbols) for test images of medium sharpness (4 c/deg) and medium contrast (12% rms). Rule 1 (dashed curves) and MIRAGE (solid curves) both predict the correct number and arrangement of features but the fit of Rule 1 is markedly better, particularly for edges. Local energy model (horizontal dotted line) captures none of the phase-dependent variation in edge and bar locations. (B) Goodness-of-fit (rms positional error) for predicted edge locations given by Rule 1 (dashed curve) and by two implementations of MIRAGE: the usual multi-channel version (solid curve) and a single channel version (dotted curve). Sudden rise in error for MIRAGE at 1 c/deg occurs because a fixed high penalty value was added to the error measure when an observed feature was not predicted.

blur varied over a large range (σ_b from 1 c/deg to 16 c/deg). Note, however, that these findings do not bear on issues about discriminability of edge location when signals are corrupted by noise.

6.5. Contour alignment—A crucial test

In this paper we addressed feature detection through two perceptual tasks—feature marking and contour alignment. To allow a more direct comparison, Fig. 10 shows the locations of bars and midpoints of the pairs of flanking edges from Experiment 1, re-plotted in the format used for Experiments 2 and 3 (Fig. 5). Though conducted on different hardware, with different tasks, in different labs, several years apart—the two sets of findings agreed remarkably closely. Phase shifts in these phase–coherent waveforms (Fig. 1) caused perceived features to translate in the direction of the phase shift, to an extent that depended on the blur, or bandwidth, of the image profile. Marked features and points of subjective alignment were offset in the same way, and to a very similar extent. Importantly, this implies that the features explicitly identified by observers in the marking task are closely related to those used in the alignment
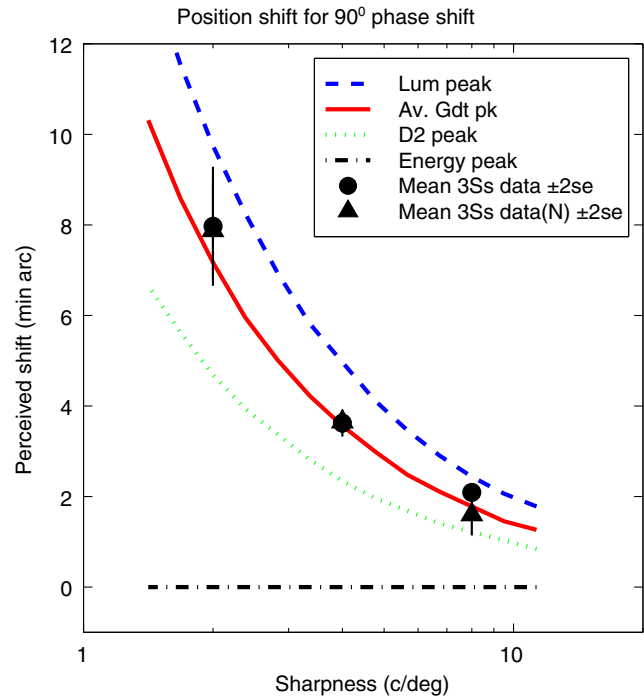

Fig. 11. Results of Experiments 2 (circles) and 3 (triangles) re-plotted as the shift in perceived position caused by a 90° phase shift in the test contour. Data are averaged over three subjects, and over light vs dark bar conditions. Error bars show ±2 s.e. ($n = 6$). Curves show the shifts expected from four different rules—see Section 6.5.

task where specific identification of feature type and polarity was not required. Fig. 11 shows that gradient peaks (edges) appear to be the most important alignment cue for these images. This figure plots the difference between perceived offsets (measured by the separation between pairs of psychometric functions; Fig. 6) for pairs of test conditions that differ by 90° of phase. This plot therefore represents the shift in perceived position induced by a 90° phase shift (e.g. from −45 to +45). The differencing operation eliminates any response bias (e.g. to select rightward responses more often than leftward) and so offers a more sensitive comparison of different potential cues. Three derivative-based cues predicted the right sort of trend, but gradient peaks (solid curve) were the most accurate. The position of the luminance peak (dashed curve) over-estimated the perceived shift, while the second-derivative peak (D2; dotted curve) under-estimated it. We can be also be fairly sure that peaks of local energy from vertical filters did not determine feature-marking or contour alignment, since the feature positions should then have been invariant with phase, but they clearly were not (Figs. 3 and 5). The control study (Experiment 3) ruled out the possibility that energy peaks from horizontal filters served to cue the perceived offsets. This cue is scrambled by inserting spatial noise between the centre and outer contours, but the observers' behaviour was unchanged. It seems
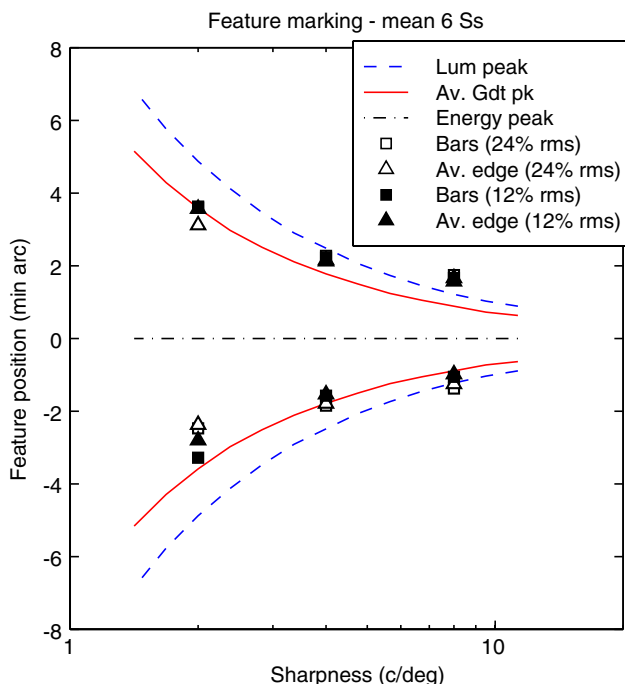

Fig. 10. Feature-marking results (Experiment 1) re-plotted in the same format as the contour alignment data (Fig. 5(C) and (D)). Group mean marked positions for bars (squares) and edges (triangles) are shown averaged across the 45°, 225° test phases (upper symbols) and across the −45°, 135° test phases (lower symbols). Results are shown separately for two contrast levels (open and filled symbols). Edge data are plotted as the mean (midpoint) position of the pair of edges that flank a bar, to enable direct comparison with the average gradient peak position (solid curve; see Fig. 4(B)).

reasonable to conclude that salient peaks of local energy in 1-D or 2-D did not control judgements of contour alignment or feature-marking.

Fig. 12(A)–(C) summarize graphically the way in which the pairs of peaks in the output of a Gaussian derivative (gradient) operator become misaligned by a 45° leftward shift in the central test contour, in a similar way to the experimentally observed shift, while Fig. 12(D)–(F) show that the output of a vertical energy operator does not shift. Fig. 12(G)–(I) remind us that energy operators can be important in other contexts: they respond well to second-order structure (here, the spatial structure of contrast modulation) that is not made explicit by linear spatial filters. But it is that sensitivity to the
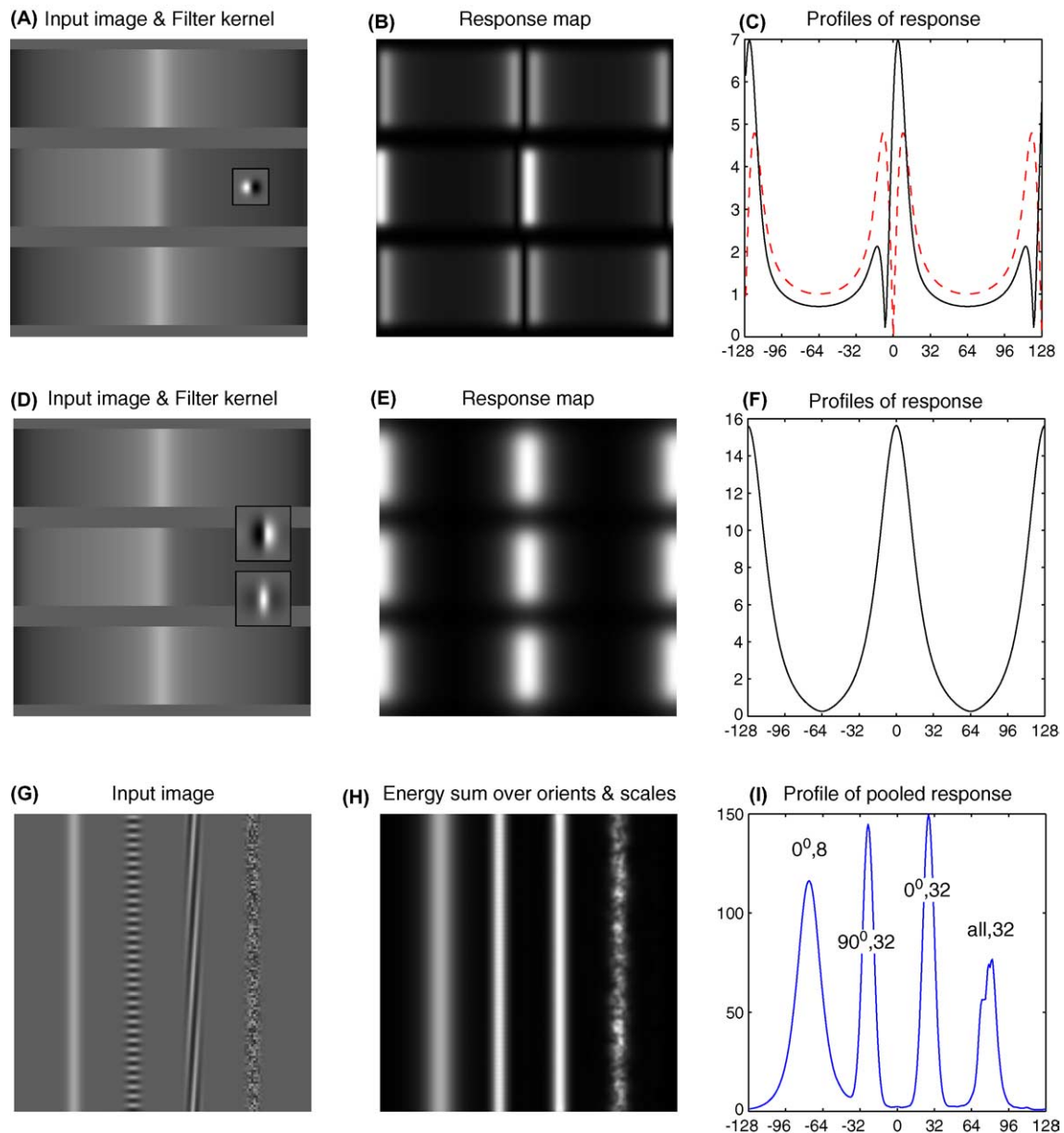

Fig. 12. Summary and comparison of the gradient- and energy-based approaches. (A) Example image from Experiment 2. Flanking contours, phase 0; centre contour, phase −45, displacement 0. Inset, Gaussian derivative operator, scale $\sigma = 4$ pixels. (B) Response image (unsigned magnitude) obtained when the smoothed gradient operator and image were convolved. (C) Response profiles taken through the upper (or lower) sections of (B) (dashed curve) and through the centre section (solid curve). Leftward displacement of the central pair of gradient peaks, accompanying the phase shift, is clear in (B) and (C). (D) Same test image as (A), but with a quadrature pair of vertically oriented, log Gabor filters (peak SF 8 c/image), used to compute the energy map shown in (E). Note that the response peaks are aligned, unaffected by phase. (G) Luminance bar (left) and three contrast-defined bars. (H) Energy responses from quadrature mechanisms (like panels (D) and (E)) but summed over four orientations (0, 45, 90, 135) and three scales (peak SF 8, 16, 32 c/image). (I) Profiles of the summed energy responses shown in (H), indicating which filter orientation and SF made the major contribution to each response peak. See Section 6.5.

contrast envelope that causes the energy profile to omit some features that people see, such as the flanking edges on bars (Fig. 3) and the local structure of the carrier in contrast-modulated sine-waves (Fig. 12(G) and (H)).

6.6. Luminance centroids

Some previous studies have suggested that the *luminance centroid* is a good predictor of perceived alignment. In an experiment directly analogous to our Experiment 2, Whitaker, McGraw, Pacey, and Barrett (1996) examined the perceived alignment of asymmetric Gaussian blobs, and found that the centroid (defined as the centre of gravity of the luminance distribution $L(x, y)$ of a blob, after subtraction of the background luminance) was a good predictor of alignment, and was rather better than the prediction made by alignment of zero-crossings (ZCs) in the second derivative $\frac{\partial^2 L}{\partial x^2}$. The asymmetry in a blob was defined by the difference in the horizontal spread coefficients (σ_1, σ_2) defining the right and left halves of the blob, and the centre and outer blobs had opposite asymmetries. The centroid rule predicted a perceived shift in alignment of $1.6 * (\sigma_1 - \sigma_2)$, while the ZC rule predicted a shift equal to $(\sigma_1 - \sigma_2)$. Data were clearly closer to the centroid rule.

However, if instead of the 1-D operator $(\frac{\partial^2 L}{\partial x^2})$ we apply a 2-D circular Laplacian operator $(\frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2})$ to these 2-D blobs, in the manner of Marr and Hildreth (1980), we find that the ZC rule (at the centre line of the blob, where $y = 0$) now predicts greater shifts, close to Whitaker et al.'s data (their Fig. 4). To see why, consider a blob with horizontal spreads σ_1 (right half) and σ_2 (left half), with fixed vertical spread (σ_1), and $\sigma_2 \leqslant \sigma_1$. With the 1-D operator (assuming its scale is small) it is simple to show (as Whitaker et al. did) that the ZCs lie at $x = -\sigma_2$ and $x = +\sigma_1$, leading to the predicted shift $(\sigma_1 - \sigma_2)$. With the 2-D operator it is straightforward to show that the ZCs (at $y = 0$) lie at $x = -\sigma_2 \sqrt{1 + \frac{\sigma_2^2}{\sigma_1^2}}$ and $x = \sigma_1 \sqrt{2}$, giving a predicted shift of $\sigma_1 \sqrt{2} - \sigma_2 \sqrt{1 + \frac{\sigma_2^2}{\sigma_1^2}}$. When plotted as a function of $(\sigma_1 - \sigma_2)$, this function lies close to the centroid rule and to the data. For example, when $(\sigma_1 - \sigma_2) = 8$ pixels, the centroid rule predicted an alignment shift of $1.6 * 8 = 12.8$ pixels, the 2-D ZC rule gave 12.9 pixels, and the observed shifts were 13–15 pixels. We confirmed the mathematical derivation of this ZC rule by checking it against a 2-D image-processing simulation in *Matlab*. The two agreed well.

This analysis shows that at least one form of edge-alignment rule—based on ZCs of the circular Laplacian operator—is quite consistent with previous data on alignment of asymmetric luminance blobs (Whitaker et al., 1996). This ZC rule therefore stands as an alternative to the centroid rule. Conversely, we can ask how well luminance centroids might account for our alignment data. If we delimit light (or dark) regions as con-

nected sets of pixels that lie above (or below) the mean luminance, and then compute their centroids, we find that a 45° phase shift produces a centroid shift of about 15.7′ for all three levels of sharpness used. This is 4–15 times larger than the perceptual shifts observed in Experiment 2. It is therefore likely that computation of luminance centroids does not contribute to perceived alignment for the class of stimuli we used.

6.7. A way forward

The two broad approaches contrasted in this paper have much in common, notably in the use of even and odd filter responses to find bars and edges respectively. But both are also incomplete. The energy model finds edges and bars correctly when phase alignment is 0° or 90°, and gives a good account of Mach Bands, but it underestimates the number of identified features, and it mislocates them when the luminance waveform is asymmetric (phase intermediate between 0 and 90). On the other hand, Rule 1 (based on luminance peaks and gradient peaks) describes the pattern of our results very well, but it fails to describe Mach Bands, where there are neither peaks of luminance nor gradient. Mach Bands occur at fairly abrupt changes in gradient, where there are peaks in the second derivative, but if we supposed that all such peaks were bars we should find far too many—flanking all edges (phase 90), where Fig. 3 clearly tells us that people do *not* see bars. We conjecture that a better account may be found by combining elements of these two approaches, and we sketch a possible avenue in Fig. 13.

The outputs of even and odd filters constitute spatial distributions of *evidence* about the presence of bars and edges respectively. A key insight from the energy model is that these two streams of evidence are not analyzed independently, but conjointly, to prevent irrelevant peaks being selected as features. Bandpass filtering necessarily introduces extra peaks and troughs into the response waveform and some of these must be rejected because they do not correspond to significant events in the image. Our results show, however, that using energy peaks as a selection rule rejects too many features, and misplaces them. Let us consider a revised selection rule: where the response of the even filter has a spatial maximum, and is greater than the odd response at that point, then we may assert a bar, and vice-versa to assert an edge. For example, Fig. 13(B) uses Gaussian first and second derivative operators as the odd and even filters. The peak and trough in the even response (dashed curve) near an edge (about $x = 100$) are not selected as bars because they do not exceed the odd response at those points. Meanwhile, similar (but larger) peaks and troughs on the Mach ramp (near $x = 400$) are accepted as bar features because they do exceed the odd response. Thus, while the energy model uses the
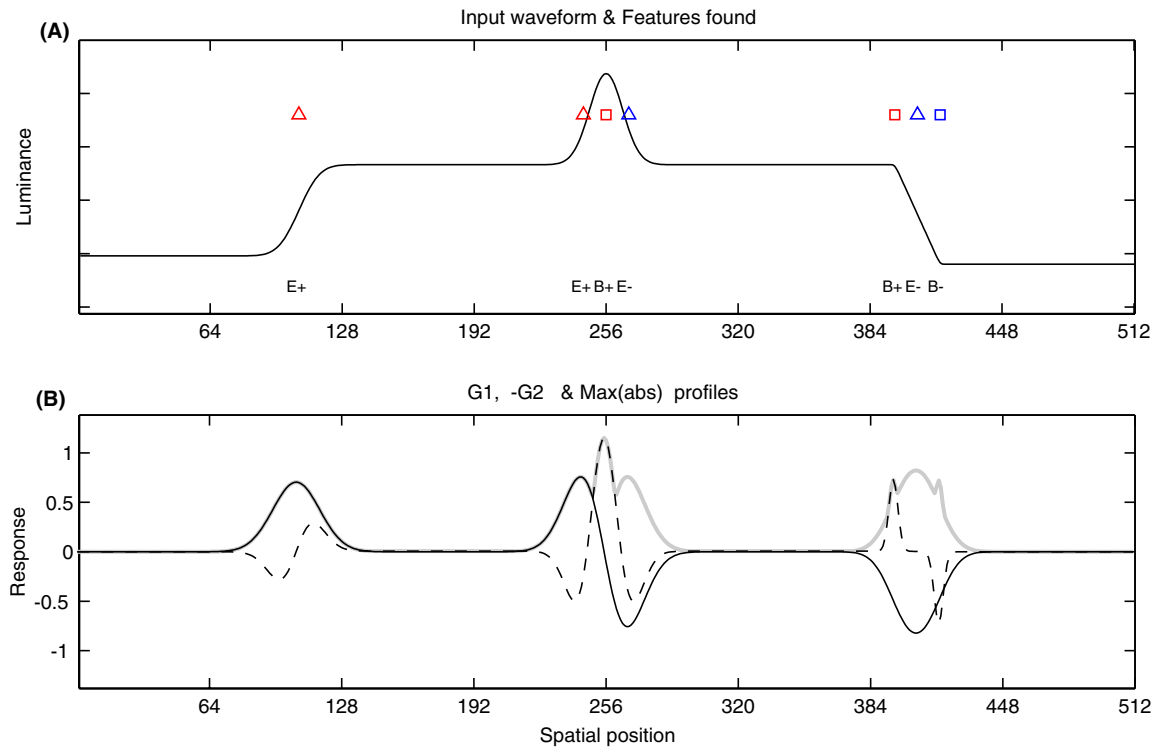
Fig. 13. How the visual system might use an envelope (max) operator to select filter response peaks as features. (A) Input waveform contains a blurred edge, a blurred bar, and a Mach ramp. (B) Thin black curve shows the output $g_1(x)$ from a Gaussian derivative (gradient) operator (scale $\sigma = 8$ pixels). Thin dashed curve shows the (inverted) output $g_2(x)$ from a Gaussian second derivative operator (scale $\sigma = 2$ pixels). Thick grey curve shows the envelope $e(x)$, where $e(x) = \max\{|g_1(x)|, |g_2(x)|\}$. Symbols in (A) represent the edges (E) and bars (B) found at spatial peaks in the envelope $e(x)$. See Section 6.6.

sum-of-squares envelope function to select feature points, we here envisage a *max* operator. The two are related via the Minkowski sum formula:

$$\text{env}(x) = \{|f_1(x)|^m + |f_2(x)|^m\}^{1/m}$$

where env(x) is the envelope function, f_1, f_2 are the even and odd filter responses and m is a parameter. When $m = 2$, env(x) is the quadratic sum used in the local energy formula, but when m is large (say, >10) env(x) effectively picks the greater of the two responses at each point. Thus by using the latter envelope function (Fig. 13(B)), we can accept some peaks in the second derivative as bars (and Mach Bands), while rejecting others where the edge-evidence is greater (Fig. 13(A)). Interestingly, for a sine-wave grating, the *max* operator finds two bars and two edges per period, which is more consistent with human perception (Georgeson, 1994, 1998) than the featureless prediction made by the quadratic sum.

Fig. 13 does no more than illustrate a possible way forward, based on a combination of ideas from the two approaches. The success or otherwise of the *max* rule will no doubt depend crucially on the type of odd and even filters adopted. A major challenge is to develop a multi-scale system based on this scheme that will work gracefully at all spatial scales, and predict features in accord with human perception. We are now developing these ideas drawn from the Gaussian-derivative scale-space literature (e.g. Lindeberg, 1998) and plan to describe them more fully in a future paper.

6.8. Conclusions

We tested two broad approaches to feature analysis in human vision—gradient-based and energy-based (see Section 1). The results were consistent with the idea that features judged in both a local and a more global task are located at peaks in the luminance profile (bars) and its first-derivative (gradient) profile (edges), but not specifically located at points of phase coherence that are peaks of local energy. Alignment of bars appeared to be determined more by the alignment of their edges than their luminance peaks. Taking a broader view, however, we have seen that both approaches are incomplete. We conjecture that computational models combining insights from the two approaches may offer more complete accounts of human feature detection.

Acknowledgments

References

Bell, A. J., & Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research, 37*, 3327–3338.

Boie, R. A., & Cox, I. J. (1987). Two-dimensional optimum edge recognition using matched and Wiener filters for machine vision. In *Proceedings of the International Conference on Computer Vision* (pp. 450–456).

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 443–446.

Burr, D. C., & Morrone, M. C. (1992). A non-linear model of feature detection. In R. B. Pinter & B. Nabet (Eds.), *Non-linear vision* (pp. 309–328). CRC Press, Inc.

Burr, D. C., & Morrone, M. C. (1994). The role of features in structuring visual images. In *Higher order processing in the visual system. Ciba foundation symposium* (184, pp. 129–141). Chichester: Wiley.

Burr, D. C., Morrone, M. C., & Spinelli, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research, 29*, 419–431.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions PAMI, 8*, 679–698.

Clark, J. J. (1989). Authenticating edges produced by zero-crossing algorithms. *IEEE Transactions PAMI, 11*, 43–57.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex: I. General characteristics and post-natal development. *Journal of Neurophysiology, 69*, 1091–1117.

Elder, J. H. (1999). Are edges incomplete? International Journal of Computer Vision*, 34*, 97–122.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*, 324–353.

Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis, 20*, 699–716.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision Research, 37*, 3367–3383.

Field, D. J., & Nachmias, J. (1984). Phase reversal discrimination. *Vision Research, 24*, 333–340.

Field, D. J., & Tolhurst, D. J. (1986). The structure and symmetry of simple-cell receptive field profiles in the cat's visual cortex. *Proceedings of the Royal Society B, 228*, 379–400.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41*, 711–724.

Georgeson, M. A. (1994). From filters to features: Location, orientation, contrast and blur. In *Higher order processing in the visual system. Ciba foundation symposium* (184, pp. 147–165). Chichester: Wiley.

Georgeson, M. A. (1998). Edge finding in human vision: a multi-stage model based on the perceived structure of plaids. *Image and Vision Computing, 16*, 389–405.

Georgeson, M. A., & Freeman, T. C. A. (1997). Perceived location of bars and edges in one-dimensional images: computational models and human vision. *Vision Research, 37*, 127–142.

Graham, N. V. S. (1989). *Visual Pattern Analyzers*. New York: Oxford University Press.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology, 148*, 574–591.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology, 160*, 106–154.

Kovesi, P. (2000). Phase congruency: a low-level image invariant. *Psychological Research, 64*, 136–148.

Kube, P., & Perona, P. (1996). Scale-space properties of quadratic feature detectors. *IEEE Transactions PAMI, 18*, 987–999.

Kulikowski, J. J., & King-Smith, P. E. (1973). Spatial arrangement of line, edge and grating detectors revealed by sub-threshold summation. *Vision Research, 13*, 1455–1478.

Lindeberg, T. (1998). Edge detection and ridge detection with automatic scale selection. *International Journal of Computer Vision, 30*, 117–154.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society B, 207*, 187–217.

Mather, G., & Morgan, M. J. (1986). Irradiation: implications for theories of edge localization. *Vision Research, 26*, 1007–1015.

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: a phase-dependent energy model. *Proceedings of the Royal Society B, 235*, 221–245.

Morrone, M. C., & Burr, D. C. (1993). A model of human feature detection based on matched filters. In P. Dario, G. Sandini, & P. Aebischer (Eds.), *Robots and biological systems: Towards a new Bionics?* (pp. 43–64). Berlin: Springer-Verlag.

Morrone, M. C., Navangione, A., & Burr, D. C. (1995). An adaptive approach to scale selection for line and edge detection. *Pattern Recognition Letters, 16*, 667–677.

Morrone, M. C., & Owens, R. A. (1987). Feature detection from local energy. *Pattern Recognition Letters, 6*, 303–313.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381*, 607–609.

Ratliff, F. (1965). *Mach Bands*. San Francisco: Holden-Day.

Ratliff, F. (1984). Why Mach Bands are not seen at the edges of a step. *Vision Research, 24*, 163–165.

Regan, D. (2000). *Human Perception of Objects*. Sunderland, MA: Sinauer Associates Inc..

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*, 455–463.

Ross, J., Holt, J. J., & Johnstone, J. R. (1981). High frequency limitations on Mach Bands. *Vision Research, 21*, 1165–1167.

Ross, J., Morrone, M. C., & Burr, D. C. (1989). The conditions under which Mach Bands are visible. *Vision Research, 29*, 699–715.

Ruderman, D. L. (1997). Origins of scaling in natural images. *Vision Research, 37*, 3385–3398.

Shapley, R. M., & Tolhurst, D. J. (1973). Edge detectors in human vision. *Journal of Physiology, 229*, 165–183.

Thomas, J. P. (1965). Threshold measurements of Mach Bands. *Journal of the Optical Society of America, 55*, 521–524.

van Deemter, J. H., & du Buf, J. M. H. (2000). Simultaneous detection of lines and edges using compound Gabor filters. *International Journal of Pattern Recognition and Artificial Intelligence, 14*, 757–777.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B, 265*, 359–366.

Watt, R. J. (1988). *Visual Processing*. Hove and London: L Erlbaum Associates.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25*, 1661–1674.

Whitaker, D., McGraw, P. V., Pacey, I., & Barrett, B. T. (1996). Centroid analysis predicts visual localization of first- and second-order stimuli. *Vision Research, 36*, 2957–2970.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness of fit. *Perception & Psychophysics, 63*, 1293–1313.

Zhang, W., & Bergholm, F. (1997). Multi-scale blur estimation and edge type classification for scene analysis. *International Journal of Computer Vision, 24*, 219–250.