1 Introduction

1.1 Background

Recent psychophysical studies have suggested that the human visual system is sensitive to variations in luminance and second-order variations in local contrast and texture. Although there is some debate about the nature of second-order vision and its relationship to first-order processing, there is now a body of results showing that they are processed separately. However, the amount, and nature, of second-order structure present in the natural environment is unclear. This is an important question because, if natural scenes contain little second-order structure in addition to first-order signals, the notion of a separate second-order system would lack ecological validity.

Two models of second-order vision were applied to a number of well-calibrated natural images. Both models consisted of a first stage of oriented spatial filters followed by a rectifying nonlinearity and then a second set of filters. The models differed in terms of the connectivity between first-stage and second-stage filters. Output images taken from the models indicate that natural images do contain useful second-order structure. Specifically, the models reveal variations in texture and features defined by such variations. Areas of high contrast (but not necessarily high luminance) are also highlighted by the models. Second-order structure—as revealed by the models—did not correlate with the first-order profile of the images, suggesting that the two types of image 'content' may be statistically independent.
defined as requiring nonlinear processing, and second-order vision is defined to be those mechanisms that provide the necessary nonlinearity. There is evidence to suggest (see later) that not only do such mechanisms exist but that they are separate from first-order processing.

1.2 Natural images
Whereas second-order vision can be demonstrated for laboratory stimuli, it is not immediately clear what, if any, second-order structure is actually present in natural scenes. The visual system is known to be relatively insensitive to second-order structure in laboratory images (Schofield and Georgeson 1999). If natural images were found to contain little second-order structure, then it might be argued that a separate mechanism for detecting second-order cues would provide the observer with too little information about the world to justify its existence. That is, the hypothesised independent mechanism for second-order vision would lack ecological validity. (Note that the mechanism might still exist as an ‘artefact’ of processing other visual cues.)

Given that second-order vision is—by definition—sensitive to variations in texture and contrast, it seems unlikely that images of everyday natural scenes would contain no second-order structure at all. Possible exceptions might include close-up images of relatively fine-grained textures, such as are found in the Brodatz (1966) collection. However, these are not ‘whole scene’ images containing multiple objects. When a scene contains multiple objects we should expect it to contain some second-order structure. This expectation has, however, not been tested. One reason for this may be that the properties of the second-order visual system are still an issue for debate (see section 1.3). Accordingly, this study represents a preliminary attempt to answer the question “what does second-order vision see in an image?” based on current ‘working hypotheses’ of how second-order vision might operate.

1.3 Second-order vision
Experiments on second-order vision typically use stimuli comprising a carrier signal (both sinusoidal gratings and white noise are commonly used carriers—see figure 1) multiplied by an envelope signal such that the contrast of the carrier at a given point is determined by the value of the envelope at that point.

Contrast modulations can be converted into luminance modulations simply by inserting a nonlinearity somewhere in the detection process. Second-order vision may employ some such nonlinearity. There are, however, many sites of nonlinearity in the visual system that could serve to demodulate images and reveal second-order signals. Early (retinal) nonlinearities (He and MacLeod 1998) would demodulate the image but integrate first-order and second-order signals prior to detection. An ability to see

Figure 1. Example second-order stimuli: (a) contrast-modulated binary white noise; (b) vertical sinusoidal carrier with a 45° contrast modulation; (c) amplitude spectrum of (b).
second-order stimuli is then only remarkable in so far as the mechanism for doing so is independent of first-order processing. Georgeson and Schofield (1998) and Schofield and Georgeson (1999) tested mixtures of luminance (first-order) and contrast (second-order) modulations of visual noise in a variety of paradigms (subthreshold summation, mixed detection, and identification at threshold) and found strong evidence for independent mechanisms. Certainly, linear summation and summation within an early nonlinearity were not supported, although summation within an energy mechanism could not be excluded. Further, Nishida et al (1997) have shown that, while stimulus-specific adaptation occurs for the detection of both first-order and second-order stimuli, cross-modal transfer of adaptation is weak and nonspecific. Thus it can be concluded that the majority of first-order and second-order stimuli are processed relatively independently in the early stages of visual processing. However, when carrier contrast is high, early nonlinearities can give rise to distortion products, which are detectable within the first-order mechanism (He and MacLeod 1998; Derrington and Ukkonen 1999; Scott-Samuel and Georgeson 1999).

First-order and second-order vision differ in their fundamental properties such as spatial and temporal resolution (see for example, Kingdom et al 1995; Lu and Sperling 1995; Smith and Ledgeway 1998; Schofield and Georgeson 1999). Here we are most interested in spatial properties. The second-order system appears to be generally less sensitive than the first-order system (Schofield and Georgeson 1999). It has a lower overall resolution than first-order vision and is spatially low-pass for contrast-modulated signals (Schofield and Georgeson 1999), but bandpass for modulations of orientation (Kingdom et al 1995). The frequency-specific motion aftereffects found by Nishida et al (1997) would suggest that the second-order system is qualitatively similar to first-order vision, being divided into frequency-selective channels. The bandwidths for channels in the two systems may not be comparable, however.

1.4 Models of second-order vision
In order to test for second-order structure in natural images it is necessary to provide a computational model that can be applied to the images. At present too little is known about second-order vision to support a definitive model; however, approximate (if somewhat hypothetical) models can be constructed as is the case here. Wilson et al (1992) propose a model for first-order and second-order vision comprising two similar mechanisms (figure 2a). First-order signals are processed by a bank of linear spatial-temporal filters. Second-order signals are processed in a separate mechanism comprising two filtering stages separated by a nonlinearity.\(^\text{(1)}\) The first stage of filtering serves to filter the carrier signal into frequency and orientation bands. The second-stage filters are the second-order analogue of the channels in the first-order system. If the first-stage filters are tuned to relatively high frequencies, then they will tend to block low-frequency first-order signals but pass the higher-frequency carrier components, preserving any contrast modulations. The nonlinearity then demodulates any second-order information ready for processing by the second-stage filters. If these are tuned to relatively low frequencies (compared to the corresponding first-stage filter), they will tend to pass any low-frequency components produced by demodulation and reject the high-frequency carrier components. Thus only second-order components pass through both stages of filtering, and since these do not normally register in the first-order mechanism, the desired independence is achieved. It will be noted that this so-called filter–rectify–filter model bears more than a passing resemblance to ‘back-pocket’ models of texture.

\(^\text{(1)}\)In this paper the terms first-order or second-order refer to the independent, parallel mechanisms for processing the two kinds of information. The two mechanisms are presumed to be further divided into spatial-frequency-specific and orientation-specific channels. The terms first-stage and second-stage filters refer to sequential stages of filtering within the second-order mechanism (see figure 2).
processing that typically employ multiple stages of filtering separated by a nonlinear mechanism (see for example Malik and Perona 1990; Landy and Bergen 1991). Such similarities serve to underline Cavanagh and Mather’s (1989) definition of second-order stimuli and suggest that modulations of texture and contrast might be detected by similar, or even the same, nonlinear mechanisms.

The precise nature of the mapping of first-stage filters onto second-stage filters is a matter of debate. A high-frequency to low-frequency mapping is preferred on the basis of maintaining mechanism independence, but is this mapping specific, with first-stage filters being wired to particular second-stage filters or more general? Sutter et al (1995) found that second-stage filters were differentially sensitive to carrier frequencies 3–4 octaves above their own preferred frequency, and thus suggest a specific mapping. In contrast, Dakin et al (1998) found that second-order channels are sensitive to all carrier frequencies higher than their own, with sensitivity increasing with carrier frequency. When the difference in preferred frequencies was low, the mapping was orientation-selective with second-stage filters being most sensitive to carrier components orthogonal to their preferred orientation. When the difference in frequency was high, the mapping was isotropic. They thus conclude that second-stage filters receive input from multiple first-stage filters and that the mapping is quite general.

Figure 2. Models of second-order vision. (a) The two-mechanism filter–rectify–filter model of motion detection as proposed by Wilson et al (1992). (b) Early pooling: one second-stage filter receives input from multiple first-stage filters. (c) Late pooling: each first-stage filter drives its own second-stage filter. Second-stage outputs can then be pooled to form second-order frequency/orientation channels.
A further issue concerns whether and at what point signals from the many channels that are implied by the filter–rectify–filter model are pooled. In a specific-mapping model, second-stage filters receive inputs from first-stage filters at many orientations. In a general-mapping model, second-stage filters also receive input across multiple frequencies. If the output of the first-stage filters were pooled prior to the second-stage (early pooling) then this pooled signal could be fed to a single second-stage filter for each second-order channel (figure 2b). If such early pooling does not take place, then multiple second-stage filters would be required for each second-order channel, each connected to a different first-stage filter (figure 2c). Second-order filter outputs might then be pooled prior to subsequent processing.

Two models were constructed in this study. The first employed a specific mapping and early pooling. This represents a computationally efficient model of second-order vision. The second model employed both a general mapping and late pooling, and was correspondingly more complex. Owing to the confound of pooling and mapping styles, it would be unwise to draw firm conclusions about preferred configurations from this study. Within each model, the bandwidths of filters in the two stages were identical and broadly consistent with those derived from psychophysical and physiological studies (see, for example, De Valois et al 1982; Snowden 1992).

2 Method

2.1 Natural images and other stimuli

Images of eight natural scenes were photographed with an SLR camera. The scenes were largely rural but did contain some man-made features including roads, paths, and occasionally buildings. Six of the eight images are shown in figures 7 and 8; the remaining images were similar to these and are excluded only to save space. A large greyscale calibration chart was included in each photograph but not included in the final images. Photographs were scanned into a computer with a flat-bed scanner. Images were then calibrated to correct for the gamma and MTF of the capture and digitisation process (for details see Thomson and Foster 1997). The final image resolution was 256 × 256 pixels with a greyscale resolution of 8 bits. The dynamic range of the calibrated images was stretched such that the full 8-bit range was utilised. Two computer-generated stimuli were used to verify the operation of the models before they were used to process the natural scenes.

2.2 Equipment and software

Models were constructed as software simulations running on a Sun SPARCstation 2 (Sun Microsystems, CA), by means of a combination of HIPS2 (SharpImage Software, New York, USA; Landy et al 1984) image-processing routines and customised software written in C. Gabor filters were used throughout and were designed to be biologically plausible to the first approximation; all had 1.0 octave frequency bandwidth and a 60° (full width at half height) orientation bandwidth. These bandwidths were used for both first-stage and second-stage filters. Full-wave rectification was used as the demodulating nonlinearity.

2.3 Specific-mapping early-pooling model

The simpler of the two models (figure 3) used a specific mapping between preferred frequencies in the two stages. First-stage filters were generated with preferred frequencies of 128, 64, 32, 16, and 8 cycles per image and preferred orientations of 0°, 30°, 60°, 90°, 120°, and 150°. These were connected to second-stage filters with preferred frequencies of 8, 4, 2, 1, and 0.5 cycles per image, respectively (that is, four octaves below the preferred frequencies of the first-stage filters). First-stage outputs were pooled across orientation before demodulation; hence the first stage of filtering was effectively isotropic. Second-stage filters had the same preferred orientations as the first-stage filters. This mapping
was specific in the sense that first-stage and second-stage frequencies were paired on a one-to-one basis.

2.4 General-mapping late-pooling model

Figure 4a shows a schematic of the more complex late-pooling model. First-stage and second-stage filtering covered the same frequency range as before, but only four orientations were used (0°, 45°, 90°, and 135°). First-stage filters were not pooled prior to demodulation. Rather, each first-stage filter was connected via its own rectifier to a bank of second-stage filters. Each first-stage filter mapped onto a unique set of second-stage filters. First-stage filters were only mapped onto second-stage filters with frequencies at least two octaves below their own. When the two preferred frequencies differed by two or three octaves, connections were made between orthogonal filters only. When the two preferred frequencies differed by more than this, all orientations were connected. Thus, at the second-stage, there were many filters with the same preferred frequency and orientation, each connected to different first-stage filters. Each second-stage filter was thus uniquely identified by its own properties and those of the underlying first-stage filter. Figure 4b shows the connectivity of the model indicating which filter pairings were implemented. In practice, the model output for a given frequency and orientation was often assessed by pooling the outputs from the many second-stage filters with the same characteristics (hence late pooling) but this is not necessary.
What does second-order vision see in an image? 1077

Figure 4. (a) Schematic showing the 8 cycles per image channel of the late-pooling general-mapping model. Second-order channels are derived from multiple first-stage frequencies, but there are limitations on the connectivity [see text and (b)]. First-stage filter outputs are not pooled before rectification so each drives its own second-stage filter. Consequently there are multiple second-stage filters for each second-order channel. (b) Table indicating the connectivity of the model; columns indicate the preferred frequency of second-stage filters, rows indicate the first-stage frequencies. An ‘x’ indicates that the filter pairing was not implemented, ‘orth-only’ indicates that only orthogonally oriented filter pairings were implemented for the given pair of frequencies, ‘all ori’ indicates that second-stage filters of all orientations were connected to first-stage filters of all orientations.
3 Results
3.1 Artificial contrast-modulated stimulus

Figure 5b shows the output of the late-pooling model in response to an artificial second-order stimulus consisting of binary noise amplitude-modulated by a sinusoid and by a raised cosine windowing function (figure 5a). The output faithfully recovers the combined modulating function which is zero around the edges of the image and sinusoidal (about a DC level) in the centre of the image. The output of the model was decomposed into separate channels from which the most strongly responding channels have been isolated and are presented here. Strong responses were recorded in the 8 cycles per image – 90° filters with the strongest response coming from that second-stage filter connected to the 64 cycles per image – 0° first-stage filter. This pattern of connectivity is unsurprising since the sinusoidal component of the modulation was vertical (90°) and only orthogonal pairings were made at this combination of frequencies. The output of this channel is shown in figure 5c; it conveys the sinusoidal component of the modulation. To prove that this filter was not simply picking out a particular frequency and orientation from the noise signal, figure 5d shows the response of an 8 cycles per image – 0° second-stage filter which gave a relatively weak response and which contains no overall structure. The coherent sinusoidal pattern in the output of the 90° filter indicates that it was genuinely responding to the contrast modulations. Other 8 cycles per image – 90° second-stage outputs gave relatively strong responses to the sinusoidal modulation. The only other channels to give a large response were the 1 cycle per image channels. The combined output of these channels is shown in figure 5e; these channels responded to the raised cosine envelope.

![Figure 5](image)

Figure 5. Processing contrast-modulated noise. (a) The input image comprising binary white noise contrast-modulated by an 8 cycles per image sinusoid and a raised cosine envelope. (b) Pooled output of the late-pooling model. (c) The output of one 8 cycles per image – 90° second-stage filter. (d) The output of a single 8 cycles per image – 0° second-stage filter. Pixel values in this image have been scaled to match the range of the image in (c). (e) Pooled outputs of all the 1 cycle per image second-stage filters.
3.2 Artificial texture modulations
Given the similarity of models of second-order vision and texture processing, it is interesting to examine the response of the former to texture-defined modulations, such as those presented by the image of figure 6a. This represents a filtered-noise carrier signal (filter centre frequency, 32 cycles per image) the orientation of which was varied sinusoidally across the image. Figure 6b shows the pooled output (from the late-pooling model) in response to this stimulus. It is just noise and does not recover the sinusoidal modulation in any detail. Inspection of the strength of response in the individual second-stage output images revealed a strong response from just a few filters. The most active second-stage filters are tabulated below; all were derived from first-stage filters sensitive to 32 cycles per image, the frequency of the carrier signal.

Figure 6. Processing texture modulations. (a) An orientation-modulated noise pattern; white noise was filtered by a variable-orientation Gabor filter with a preferred frequency of 32 cycles per image; the orientation of the filter varied sinusoidally with horizontal position. (b) Pooled output of all the second-stage filters in the late-pooling model. (c) and (d) Outputs of two 2 cycles per image $^\circ 90$ second-stage filters; the filter for (c) was connected to a 32 cycles per image $^\circ 135$ first-stage filter, the first-stage filter for (d) was tuned to 32 cycles per image and $45^\circ$. (e) Output of the 4 cycles per image $^\circ 90$ second-stage, 32 cycles per image $^\circ 0$ first-stage filter. (f), (g), and (h) Outputs of other active filters with properties given in table 1.
The output images for these channels are shown in figures 6c–6h. The most active pair of filters (figures 6c and 6d) are both 2 cycles per image and vertical at the second stage, which corresponds to the modulation frequency and orientation. The modulating signal is clearly recovered in these images; however, the two respond to different, orthogonal, carrier orientations and hence produce output responses that are the negative of each other. Information is thus lost when the two filter outputs are simply pooled. This information can be preserved, however, if the filter outputs are kept separate or combined in an opponent mechanism (cf Landy and Bergen 1991). The other active filters might be regarded as spurious, except the 4 cycles per image filter (figure 6e), which, being connected to a horizontal first-stage filter, signals the points in the image where the carrier orientation passes through horizontal. In effect, this image conveys the zero crossings of the texture pattern—the orientation of the texture varies sinusoidally from $-45^\circ$ to $+45^\circ$, and thus passes through $0^\circ$ twice per cycle of the modulating signal. The flexibility of having filters with the same second-stage properties but different first-stage properties is limited to late-pooling models and is crucial for the identification of the modulating signal in this example. In the early-pooling model, responses arising from all carrier orientations would be averaged before rectification and thus the modulation in this image would not be detected.

### 3.3 Natural images

The two models were each applied to all of the eight natural images. Six of these images, together with the output images from the early-pooling model are shown in figures 7 and 8 (the remaining two images produced very similar results). Two of the images are shown in figures 7a and 7d. Figure 7b shows the output of the early-pooling model in response to the image of figure 7a. Figure 7c shows the response of the late-pooling model to the same image. As was the case with all the images, the two models produced very similar results when all the second-order channels were pooled. Figure 7e is the output of the early-pooling model in response to the image of figure 7d. The images of figure 8 are paired as described in the legend. The models clearly respond strongly in areas of high image contrast, such as the edges of the road in figure 7a and the area around the car in the distance. Textured areas (for example the grass verge in figure 7a) also yield a strong response. Smooth areas, such as the road in figure 7a, produce a low response despite having high luminance. The differential response to image textures is illustrated well in figures 7d and 7e (arrows). Areas of grass, in figure 7d, with a relatively coarse grain produce a higher response in figure 7e than do areas with a relatively fine grain. Similar second-order structure can be observed in the images of figure 8.

At first glance, some of the output images may look like low-pass filtered versions of the input images, or the negatives of such images; this, however, was not the case as was proven by the correlational analysis presented below.
In an attempt to quantify the relationship between first-order and second-order structure in an ensemble of natural images, the results of processing the eight scenes were correlated with the first-order (that is intensity) profile of the original images. For comparison, the second-order responses were also correlated with the rms contrast profiles of the original images. The values of 100 randomly chosen pixels from each of the eight output images (pooled model) were correlated with the spatially averaged local intensity and rms contrast at matching points in the input images. Intensity values were taken as the average of the pixel values in a 10 by 10 square centred on the pixel in question; rms values were calculated as the standard deviation divided by the mean within a similar area. The overall correlation between the second-order response and local rms contrast (taken over all eight images) was 0.34 (significant at $p < 0.05$) while that between second-order response and intensity was $-0.05$ (not significant). Further, while individual images sometimes yielded significant correlation between intensity and the second-order response, the sign of such correlations varied from image to image indicating an unstable relationship. The relationship between second-order response and contrast varied in strength but was always positive. It is interesting to note that rms contrast as defined here is itself nonlinear with respect to image

![Figure 7. Processing natural images. (a) and (d) Two natural images (dynamic range adjusted for presentation). (b) and (c) The pooled response of the early-pooling and late-pooling models to image (a). (e) The pooled response of the late-pooling model to image (d).](image-url)
intensity (it varies as the reciprocal of intensity). It is therefore negatively correlated with intensity. Thus, whereas second-order structure and rms contrast are clearly related, they differ in one key respect—the former is (over an ensemble of images) independent of intensity, whereas the latter is not. This suggests that the division of images into first-order and second-order structure is a more efficient coding than a division based on intensity and contrast variations.

Figure 8. (a)–(d) Four natural images and their second-order structure. (e)–(h) As produced by the early-pooling model.
The pooled second-stage response is not the only second-order signal that might be of use to higher processing areas within the visual system. In principle, all of the second-order channels could be processed independently. In the case of the late-pooling model, analysis might even be done on the basis of discrete first-stage and second-stage filter pairs. For illustration, only some example output images are shown here. The images are based on figure 7a (chosen because it is a relatively busy image), but other images in the set produced similar results. Figure 9 shows second-stage outputs for the early-pooling model pooled across orientation but kept separate in frequency. The lowest frequency channels shown (1 and 2 cycles per image) divide the image on the basis of texture (that is rough versus smooth regions). Higher-frequency channels are rather more concerned with fine detail (as might be expected), such as the edges of the road and the location of the car.

Figure 9. Frequency bands. (a) Input image. (b)–(e) Second-stage outputs of the early-pooling model further pooled across orientation; preferred frequencies were 1, 2, 4, and 8 cycles per image respectively. The 0.5 cycle per image channel, not shown, produced a small DC component only.

Figure 10 shows the output of the individual, oriented, 8 cycles per image second-order channels. As might be expected, the car produces a response at all orientations, whereas other features, such as the edges of the road, produce responses at specific orientations only.

The complete response of the late-pooling model to the image of figure 7a is documented on the Perception website (http://www.perceptionweb.com/perc0900/schofield.html) and on the CD ROM archive accompanying issue 12 of the journal. Thus, the interested reader can inspect the output image for any second-stage filter.
Images like those of figures 7, 8, 9, and 10 suggest that sensitivity to second-order modulations could provide the visual system with an enhanced representation of natural scenes, compared with the one that would be available to purely linear (first-order) mechanisms. That is, natural images contain features which would preferentially stimulate second-order mechanisms if such mechanisms are indeed available. Thus, it can be concluded that second-order vision has ecological validity. The nature of the information carried by the second-order components of an image has been demonstrated but not formally quantified. Some changes in the second-order profile of an image may be coincident with changes in first-order properties (for example, changes in surface texture may be associated with changes in luminance). Over an ensemble of images, however, second-order structure was found to be independent of first-order structure.

**Figure 10.** Orientation bands. (a) Input image. (b)–(g) Second-stage outputs of the 8 cycles per image channel of the early-pooling model. The preferred orientations of the second-stage filters were 0°, 30°, 60°, 90°, 120°, and 150°, respectively.
The models of second-order vision presented here are remarkably similar to those suggested for texture segmentation. When second-order information is pooled across orientation channels (as might be most appropriate for the detection of contrast modulations), information conveyed by modulations of orientation is lost. It is well known that the visual system is sensitive to such modulations of texture, as is illustrated by the vertical, low-frequency, wave percept of figure 6. This loss of information can be avoided if orientation channels are combined in opposition or processed separately as is typically the case in models of texture segmentation (Landy and Bergen 1991). This would suggest that early pooling is an inappropriate model for texture processing. Even in the case of late pooling, human sensitivity to orientation changes would suggest that the pooling process is flexible enough not to pool in all cases.

The limited nature of the models used in this study means that it would be unwise to favour one type of mapping (general or specific) over the other on the basis of this study. After pooling, the two models produced nearly identical output images. Further modelling, combining late pooling with the two kinds of mapping, might be used to support one model over another, especially if the results can be related to psychophysical findings.

Acknowledgements. This work was supported by a BBSRC project grant (S03969) awarded to Mark Georgeson who is also thanked for many helpful discussions on the topic. I thank Mitch Thomson for the kind provision of the calibrated natural images used in this study, Steve Guest for his careful proofreading, and the two anonymous reviewers for their helpful comments.

References

Brodatz P, 1966 Textures: A Photographic Album for Artists and Designers (New York: Dover)
Dakin S C, Mareshal I, Hess R F, 1998 “Sensitivity to amplitude modulation depends on carrier spatial frequency and orientation” Perception 27 Supplement, 38
Harris L R, Smith A T, 1992 “Motion defined exclusively by second-order characteristics does not evoke optokinetic nystagmus” Visual Neuroscience 9 565 – 570
He S, MacLeod D I A, 1998 “Contrast-modulated flicker: dynamics and spatial resolution of the light adaptation process” Vision Research 38 985 – 1000
Smith A T, Ledgeway T, 1998 “Sensitivity to second-order motion as a function of temporal frequency and eccentricity” Vision Research 38 403 – 410