representation of the scene that contains important images tokens, such as edges and other basic image features, and that this representation permits further analysis of the scene to be done more efficiently by the brain. This concept has greatly influenced much of the research done in computer vision.

A number of variations of second derivative operators have been devised in various attempts to overcome their deficiencies. Some examples of this include the work of Fleck [22], Haralick [32], and Sarkar and Boyer [83]. Fleck and Haralick used directional second derivatives to reduce the influence of noise, with Fleck also employing first and third derivative information to eliminate the detection of false positives. Sarkar and Boyer adopted the optimality criteria proposed by Canny [11, 12] to develop infinite impulse response filters for the detection of edges via zero crossings.

Canny [11, 12] formalized the problem of the detection of step edges in terms of three criteria: good detection; good localization; and uniqueness of the response to a single feature. Subsequently Spacek [87] and Deriche [16] followed Canny's approach to develop similar operators; Deriche allowing the operator to have an infinite impulse response and Spacek modifying the response uniqueness criterion. An objection to these 'optimal' detectors is that they are only optimal in a very limited domain, that of one dimensional step edges in the presence of noise. At 2D features such as corners and junctions where the intensity gradient becomes poorly defined these detectors have difficulties.

Thus, a major problem with gradient based operators is that they use a single model of an edge, that is, they assume edges are step discontinuities. In an ideal system a feature detector would mark features wherever a good artist would draw features when making a sketch of a scene. An artist produces marks in a sketch for a wide range of feature types, not just step edges. Marks are drawn to indicate line, roof and step edges along with other features such as shadow boundaries, highlights, and presumably a range of other (unknown) feature types. Perona and Malik [68] point out that many image features are represented by some combination of step, delta, roof and ramp profiles. For example, a very commonly encountered feature type is the occluding boundary of a convex object, such as a ball. If the ball surface

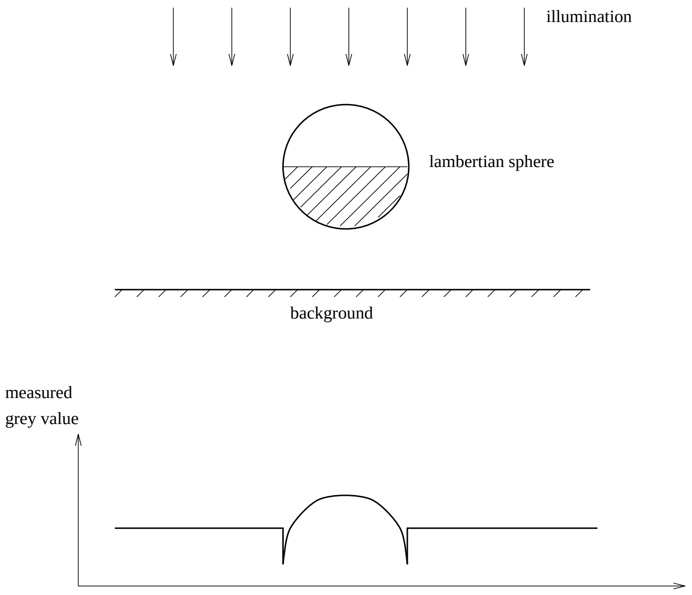


Figure 1: Intensity profile observed across a lambertian sphere against a plain backgound with overhead illumination. The occlusion boundary is not a simple step edge.

is lambertian and the illumination is aligned with the viewing direction the feature profile will consist of an intensity profile that starts off brightest at the mid-point of the ball and then gets darker as our view moves across the ball as a result of the surface normal becoming perpendicular to our viewing direction, and finally culminating in a step jump to the grey level of the background (Figure 1).

In this simple, idealized situation we have a feature that is considerably more complex than a step edge. In practice the situation will be far more awkward; the ball surface is unlikely to be lambertian, lighting can be from any direction, there may be mutual illumination effects between the ball and other objects, and of course, the background may not be uniform. For this reason the word 'feature' will be generally used in this thesis rather than the word 'edge' in order to emphasize the aim of finding all important features that represent points of high information content, not just step edges. The definition of what a 'feature' is will be deliberately left vague, though subsequent sections which describe the phase congruency model of feature perception will offer a possible definition.

Some might argue that an automated feature detector does not need to attempt to emulate human sketching skills. However, the interest in producing feature detectors has been primarily inspired from the ability of artists to produce line drawings[2]. Artists have shown us that line drawings can provide very compact yet effective descriptions of scenes. Indeed, in the assessment of any automated feature detector perhaps the best we can do is to compare its output against a line drawing of the same scene made by an expert reproductive artist. After all it is artists who are our best experts in representing scenes via line drawings. It is probably fair to say that excessive emphasis has been placed on finding 'optimal' step edge detectors and the original objective, that of finding points of high information content in images has been forgotten. Just because a detector is effective in finding and localizing noisy step edges in a scene does not mean that it will represent the information in the scene well.

A second problem with gradient based edge detectors is that they typically characterize edge strength by the magnitude of the intensity gradient. Thus the perceived strength or significance of an edge is sensitive to illumination and spatial magnification variations. Intensity gradient has units of *lux/radian* (pixel coordinates represent viewing direction and hence have angular units)[3]. Intensity gradients in images depend on many factors, including scene illumination, blurring and magnification. For example, doubling the size of an image while leaving its intensity values unchanged will halve all the gradients in the image. Any gradient based edge detection process will need to use a threshold modified appropriately. However, in general, one does not know in advance the level of contrast present in an image or its magnification. The image gradient values that correspond to significant edges are usually determined empirically.

---

[2] Here the distinction is made between line drawings, which contain only lines, and sketches which may also include shading.

[3] Strictly speaking, image grey values should not be called 'intensity values'. Intensity is defined as the luminous flux that is emitted per solid angle and is a property that is associated with a light source. Intensity has units *candelas* ($lumens/steradian$). In constructing an image a camera measures the illumination at each point in the image plane that is received from a scene. Thus, image grey values have units *lux* ($lumens/m^2$). Despite this, the use of the term 'intensity value' for an image grey value appears to be commonly accepted. David Marr used the term in this manner in his book [54].

Little guidance is available for the setting of thresholds, indeed Faugeras[4] can only offer the following advice:

> "Thresholding is a plague that occurs in many areas in engineering, but to our knowledge it is unavoidable and must be tackled with courage".

A limited number of efforts have been made to determine threshold values automatically. In his thesis, Canny [11] sets his thresholds on the basis of local estimates of image noise obtained via Weiner filtering. However, the details of setting thresholds on this basis, and the effectiveness of this approach are not reported. Canny also introduced the idea of thresholding hysteresis which has proved to be a useful heuristic for maintaining the continuity of thresholded edges, though one then has the problem of determining two threshold levels. Sarkar and Boyer [83] also employed Weiner filtering to estimate the derivative of the noise output in their zero crossing based detector. Having an estimated slope of the noise response allowed them to set thresholds appropriately. However, this process required them to take three more derivatives *after* the image had been filtered by their edge operator. This presumably limited the quality of the estimate of the derivative of the noise output.

Kundu and Pal [50] devised a method of thresholding based on human psychophysical data where contrast sensitivity varies with overall illumination levels. However, it is hard to provide any concrete guide to the fitting of a model of contrast sensitivity relative to a digitized grey scale of 0–255. More recently Fleck [24, 23] suggested setting thresholds at some multiple (typically 3 to 5) of the expected standard deviation of the operator output when applied to camera noise. This approach of course, requires detailed a priori knowledge of the noise characteristics of any camera used to take an image. Noise is always a concern for gradient based detectors. The main tool used to reduce the influence of noise is spatial smoothing. However, smoothing degrades feature localization, and 2D feature positions such as corners can be severely corrupted (see Perona and Malik [69]). With high degrees of smoothing feature locations can move significantly, and distinct features may

---

[4] Olivier Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* MIT Press 1993, p117.

merge. It is very unsatisfactory for the perceived location of a feature to depend on how much smoothing was required to overcome the influence of noise. This issue will be considered in more detail in the next chapter.

Bergholm [5] adopts the scale-space model in developing his *edge focusing* approach to edge detection, and in doing so addresses an number of problems associated with gradient based detectors. He observes that to eliminate the influence of noise on a gradient based detector a heavily smoothed image is required, but this degrades edge localization. To achieve good localization no smoothing should be used but then noise becomes a problem. Bergholm's solution is to start with an edge map at a heavily smoothed scale. He then proceeds to calculate an edge map at a slightly finer scale *but only at pixels in the image connected to edge pixels found at the previous scale.* The old edge points are discarded, the new ones at the slightly finer scale retained, and the process is repeated. In this manner edges are propagated out from their initial, rough locations and focused to their correct positions at the finest scale. An important point is that the problem of noise is overcome by starting with edges at a coarse scale and only looking for edges in adjacent pixels as scale is gradually reduced. Another attractive feature is that edge thresholding is only required to generate the initial edge map. However if this initial map is incorrectly thresholded at too high a level then many features will never be found. Conversely, if the threshold is too low many noise features will be found and these will be propagated down to the finest scale.

The discussion so far has been directed at gradient based detectors though, of course, other types of detectors have been developed. For example the *weak membrane* approach of Blake and Zisserman [6] involves minimizing a global energy function over the image in order to solve for a surface function that fits the image in a manner that is considered to be appropriate. Blake and Zisserman's energy measure is a weighted combination of terms representing the deviation of the surface function from the image, the square of the slope of the function, and the contour length of the function. This can be interpreted as fitting a weak membrane to the image data in such a way that discontinuities are preserved. An objection to this approach is that the energy term is not dimensionally consistent with different types of quantities being added together. This makes the result very sensitive to the relative weightings of the terms that make up the energy.

Noble [64] devised a number of grey level morphological operations to detect edges. She develops a *dilation-erosion residue* operator which is analogous to a first derivative operator and is used as an edge strength map. A second operator called the *signed maximum dilation-erosion residue* (analogous to a second derivative operator) is used to guide the tracing of edges, and to classify the responses to the dilation-erosion residue operator. While Noble's approach is morphological, the steps involved can be interpreted in terms of differential operators. Thus it depends on using a simple edge model and it does not escape the thresholding problem.

Perona and Malik [69] devised an approach to edge detection using anisotropic diffusion. They developed an approach to scale space smoothing that is based on the heat diffusion equation. To detect edges they make the conduction coefficient a function of the image gradient to impede the flow of 'heat'. Thus step discontinuities in the image form local barriers to the diffusion process. Over repeated iterations of the diffusion process step edges in the image become sharper and regions between the step discontinuities become smoother. Final extraction of the edges then becomes straightforward. A very significant attribute of this approach is that feature positions remain stable over scale. All that changes with scale is the level of contrast (heat difference) required for a feature to persist. However, this approach only detects step edges and is very much dependent on local image contrast.

Another interesting approach that has been developed recently by Smith and Brady [85] is the SUSAN edge finder. This non-linear technique involves indexing a circular mask over the image and at each location determining the area of the mask having similar intensity values to the centre pixel value. This segment of the mask is denoted the Univalue Segment Assimilating Nucleus (USAN). Locations in the image where the USAN is locally at a minimum (locally the Smallest USAN, hence SUSAN) mark the positions of step and line features. The detector performs well, and its tolerance to noise is a significant attribute. However, the detector is not invariant to image contrast as it requires the setting of a threshold which is used to decide whether or not elements of the mask are 'similar' to the centre value when determining the size of the USAN. This threshold specifies the minimum edge contrast that can be detected.

The discussion above represents a generalized overview and sampling of existing edge detection techniques. Others have conducted far more comprehensive reviews (for example Noble [64]), and it is not intended to repeat such a review here. The main purpose of this overview is to point out that almost all existing edge detectors are based on the calculation of intensity gradients or some other measure of the spatial variation of intensity across the image. These measures are dimensional quantities and hence depend on image contrast and spatial magnification. Thus the fundamental problem is that one does not know in advance what level of edge strength corresponds to a significant feature. As a result, edge thresholds are generally set by humans viewing the output and adjusting the threshold until the result is deemed acceptable. This is *not* automated feature detection.

2.3 Local energy and phase congruency

The local energy model of feature perception is a relatively new model. It is not based on the use of local intensity gradients for feature detection. Instead it postulates that features are perceived at points in an image where the Fourier components are maximally in phase[5]. For example, when one looks at the Fourier series that makes up a square wave all the Fourier components are sine waves that are exactly in phase at the point of the step at an angle of 0 or 180 degrees depending on whether the step is upward or downward. At all other points in the square wave individual phase values vary, making *phase congruency* low. Similarly one finds that phase congruency is a maximum at the peaks of a triangular wave (at an angle of 90 or 270 degrees). A particularly important point about using phase congruency to mark features of interest is that one is not making any assumption about the

---

[5] It should be emphasized that when phase is referred to in this thesis it is *local phase* that is being considered. That is, we are concerned with the local phase of the signal at some position x. This is distinct from phase values that one might obtain, say, from a FFT of a signal in which phase values will be the phase offsets of each of the sinusoidal basis functions in the decomposition.

shape of the waveform at all. One is simply looking for points in the image where there is a high degree of order in the Fourier domain.

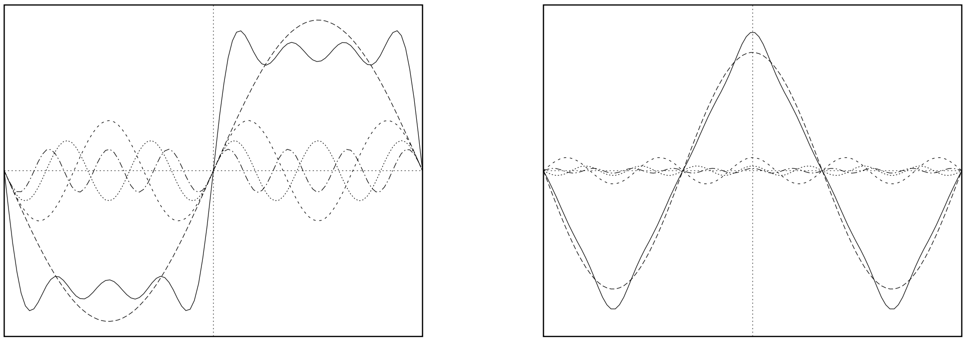


Figure 2: Construction of square and triangular waveforms from their Fourier series. In both diagrams the first few terms of the respective Fourier series are plotted with broken lines, the sum of these terms is the solid line. Notice how the Fourier components are all in phase at the point of the step in the square wave, and at the peaks and troughs of the triangular wave.

A wide range of feature types give rise to points of high phase congruency. These include step edges, line and roof edges, and Mach bands. It was, in fact, investigations into the phenomenon of Mach bands by Morrone et al. [62] that led to the development of the local energy model. Mach bands are illusory bright and dark bands that appear on the edges of trapezoidal intensity gradient ramps, for example, on the edges of shadows. The classical explanation for the perception of Mach bands has been lateral inhibition (see Ratliff [74]). However, this explanation fails in that it predicts maximal perception of Mach bands on step edges, where in fact we see none. In their paper, Morrone et al. show that at the points where we perceive Mach bands the Fourier components of the signal are maximally in phase (though not exactly in phase); this led to their hypothesis that we perceive features in images at points of high phase congruency. Further work by Morrone and Burr [60] and Ross et al. [80] went on to show that this model successfully explains a number of other psychophysical effects in human feature perception. Other studies of the sensitivity of the human visual system to phase information include that by Burr [8], Field and Nachmias [21] and du Buf [18]. Fleet [25] argues strongly for the use of phase information in the calculation of image velocities. He shows that the motion of contours of constant phase in images provide a better measure of the motion field than contours of constant intensity amplitude in the image. Phase information is more robust to noise, and shading and contrast variations in the image.

The classic demonstration of the importance of phase was devised by Oppenheim and Lim [65]. They took the Fourier transforms of two images and used the phase information from one image and the magnitude information of the other to construct a new, synthetic Fourier transform which was then back-transformed to produce a new image. The features seen in such an image, while somewhat scrambled, clearly correspond to those in the image from which the phase data was obtained. Little evidence, if any, from the other image can be perceived. A demonstration of this is repeated here in Figure 3.

With phase data demonstrated as being so important in the perception of images it is natural that one should pursue the development of a feature detector that operates on the basis of phase information. From their work on Mach bands Morrone and Owens [61] quickly recognized that the local energy model had applications in feature detection for computer vision.

2.3.1 Defining phase congruency

We shall first consider one dimensional signals. The phase congruency function is developed from the Fourier series expansion of a signal, I at some location, x,

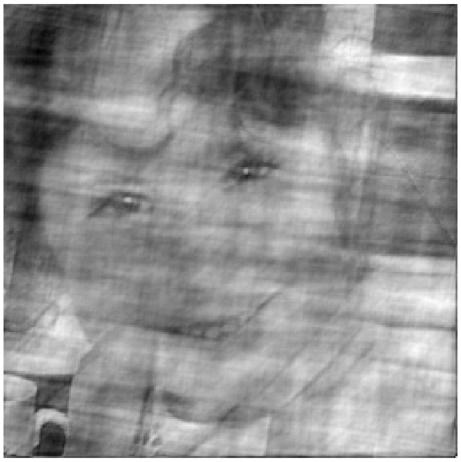
$$\begin{aligned} I(x) &= \sum_n A_n \cos(n\omega x + \phi_{n_0}) && (1) \\ &= \sum_n A_n \cos(\phi_n(x)) \ , && (2) \end{aligned}$$

where A_n represents the amplitude of the n^{th} cosine component, ω is a constant (usually 2π), and ϕ_{n_0} is the phase offset of the n^{th} component (the phase offset also allows sine terms in the series to be represented). The function $\phi_n(x)$ represents the *local* phase of the Fourier component at position x.

Morrone and Owens define the phase congruency function as

$$PC(x) = max_{\overline{\phi}(x)\in[0,2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) - \overline{\phi}(x))}{\sum_n A_n} \ . \quad (3)$$

(a) image providing magnitude data (b) image providing phase data



(c) phase and amplitude mixed image

Figure 3: When phase information from one image is combined with magnitude information of another it is phase information that prevails.

The value of $\overline{\phi}(x)$ that maximizes Equation 3 is the amplitude weighted mean local phase angle of all the Fourier terms at the point being considered. Taking the cosine of the difference between the actual phase angle of a frequency component and this weighted mean, $\overline{\phi}(x)$, generates a quantity approximately equal to one minus half this difference squared (the Taylor expansion of $\cos(x) \approx 1 - x^2/2$ for small x). Thus finding where phase congruency is a maximum is approximately equivalent to finding where the weighted variance of local phase angles, relative to the weighted average local phase, is a minimum (see Figure 4).

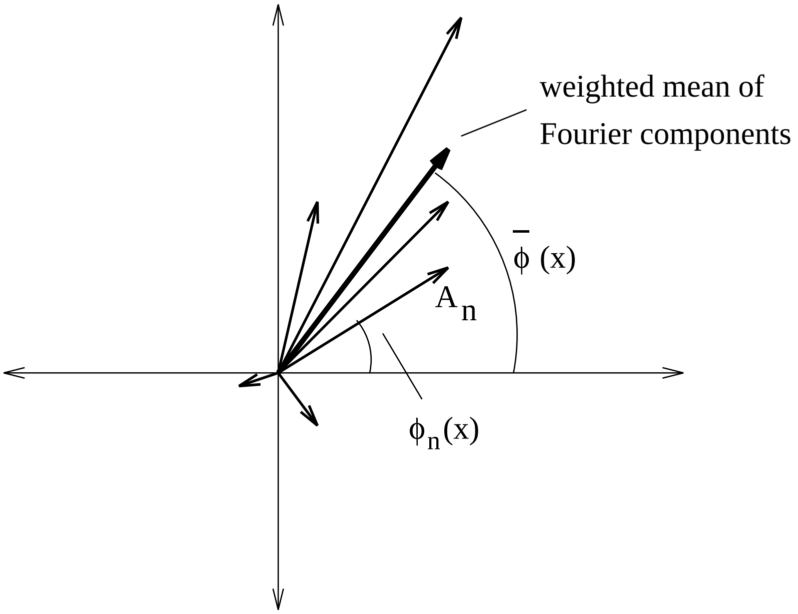


Figure 4: Polar diagram of the components of a Fourier series at a point in a signal. The series is represented as a sequence of vectors, each vector having a length A_n and local phase angle ϕ_n.

2.3.2 Local energy

As it stands phase congruency is a rather awkward quantity to calculate. As an alternative to this Venkatesh and Owens [89] show that points of maximum phase congruency can be calculated equivalently by searching for peaks in the local energy function. The local energy function is defined for a one dimensional luminance profile, $I(x)$, as the modulus of a complex number,

$$E(x) = \sqrt{I^2(x) + H^2(x)}\ , \tag{4}$$

where the real component is represented by $I(x)$ and the imaginary component by $iH(x)$, where $i = \sqrt{-1}$ and $H(x)$ is the Hilbert transform of $I(x)$ (a 90 degree phase shift of $I(x)$).

Venkatesh and Owens prove that energy is equal to phase congruency scaled by the sum of the Fourier amplitudes, that is

$$E(x) = PC(x)\sum_n A_n. \tag{5}$$

Thus the local energy function is directly proportional to the phase congruency function, so peaks in local energy will correspond to peaks in phase congruency.

Venkatesh and Owens' formal proof is not repeated here but the relationship between phase congruency, energy and the sum of the Fourier amplitudes can be seen geometrically in Figure 5. The local Fourier components are plotted as complex vectors adding head to tail. The sum of these components projected onto the real axis represent $I(x)$, the original signal, and the projection onto the imaginary axis represents $H(x)$, the Hilbert transform. The magnitude of the vector from the origin to the end point is the total energy, $E(x)$. One can see that $E(x)$ is equal to $\sum_n A_n cos(\phi_n(x) - \overline{\phi}(x))$. Recalling that phase congruency is equal to $\sum_n A_n(x) cos(\phi_n(x) - \overline{\phi}(x)) / \sum_n A_n$ we can see that phase congruency is the ratio of $E(x)$ to the overall path length taken by the local Fourier components in reaching the end point. Thus, one can clearly see that the degree of phase congruency is independent of the overall magnitude of the signal. This provides invariance to variations in image illumination and/or contrast.

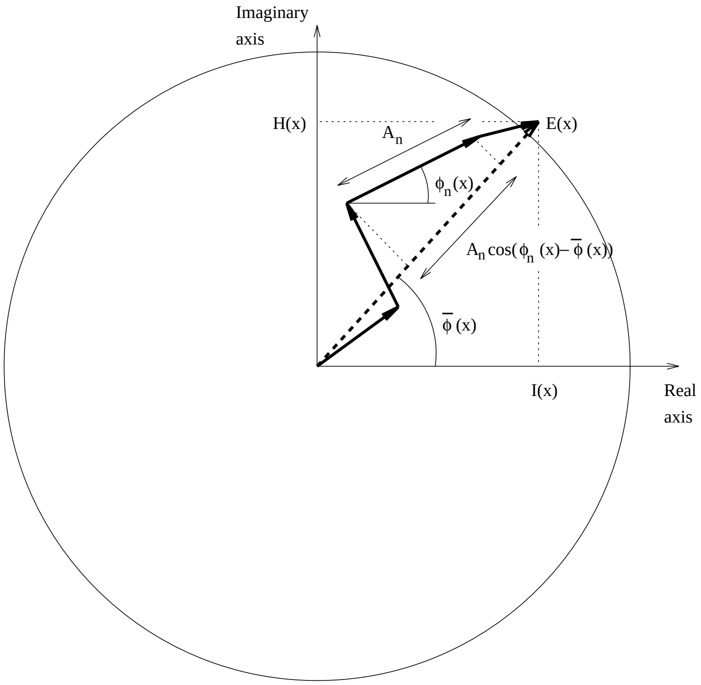


Figure 5: Polar diagram showing the Fourier components at a location in the signal plotted head to tail. This arrangement illustrates the construction of energy, the sum of the Fourier amplitudes and phase congruency from the Fourier components of a signal.

Rather than compute local energy via the Hilbert transform of the original luminance profile one can calculate a measure of local energy by convolving the signal with a pair of filters in quadrature. The signal is first convolved with a filter designed to remove the DC component from the image. This result is saved and the image is then convolved with a second filter that is in quadrature with the first (the Hilbert transform of the first). This gives us two signals, each being a band passed version of the original, and one being a 90 degree phase shift of the other. The results of the two convolutions are then squared and summed to produce a local energy function. Odd and even-symmetric Gabor functions can be used for the quadrature pair of filters. Thus local energy is defined by

$$E(x) = \sqrt{(I(x) * M^e)^2 + (I(x) * M^o)^2} \, , \tag{6}$$

where M^e and M^o denote the even and odd symmetric filters in quadrature. Figure 6 illustrates the calculation of local energy on a synthetic signal containing a variety of features.

The calculation of energy from spatial filters in quadrature pairs has been central to many models of human visual perception, for example those proposed by Heeger [33, 34, 36], Adelson and Bergen [1] and Watson and Ahumada [93] to name just a few. The significance of Venkatesh and Owens' work is that they provide another explanation for the perceptual importance of energy: Peaks in the energy function correspond to points where phase congruency is a maximum.

From this early work by Morrone et al. [62], Morrone and Owens [61] and Venkatesh and Owens [89] the local energy model was developed further. Owens et al. [67] investigated the idempotency properties of the local energy feature detector. They argue that when any feature detecting operator is applied to its own output it should not change the output. That is, the primal sketch of a primal sketch should be itself. Gradient based detectors fail in this respect because they attempt to mark 'edges' on each side of any line feature in an image. Local energy, on the other hand, produces a single response on a line feature, and hence satisfies the idempotency requirement. Venkatesh and Owens [88] investigated the classification of image features via the phase angle at which phase congruency occurs. In this manner they show how step, line and shadow edges can be distinguished from each other.

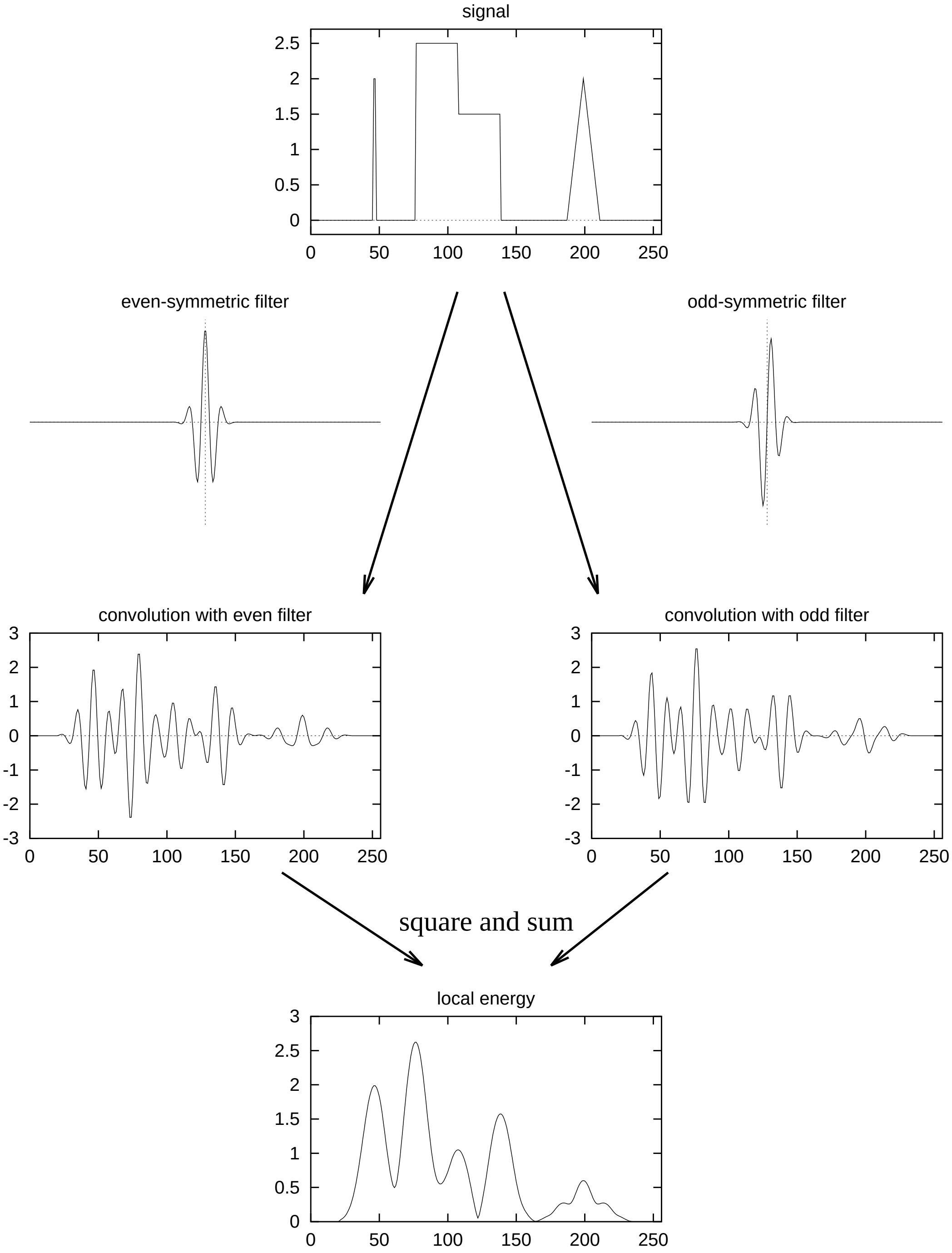


Figure 6: Calculation of local energy via convolution with two filters in quadrature.

Aw et al. [4] in their work on image compression make use of the fact that local energy makes no assumptions about the intensity profiles of features. They used local energy to detect features across a range of images, collecting information about commonly occuring intensity profiles of features in images. This catalogue of feature profiles enabled them to efficiently encode images for compression.

Owens [66] identifies the conditions under which images have no local maxima in local energy, and hence are feature free. She also investigates image transformations under which image features are preserved. It is pointed out that some image operations, such as addition between images, can destroy or create image features. She proposes two new operators for the interaction between images which do not corrupt feature structures within images. These operators are analogous to complex multiplication and complex division. Using these operators Owens shows how it is possible to decompose a signal into its feature component and its feature-free component.

Other researchers who have studied the use of local energy for feature detection are Perona and Malik [68], Freeman [29] and Ronse [78]. Perona and Malik's work on local energy is interesting in that they arrive at a generalization of the model without using the concept of phase congruency. They point out that image features are generally composed of combinations of step, delta, roof and ramp structures. Under these conditions it is shown that linear filters will produce systematic errors in localization. Perona and Malik go on to show that a quadratic filtering approach results in the correct detection and localization of composite features. That is, instead of looking for maxima in $(I(x) * M)$ one should look for maxima in $\sum_i (I(x) * M_i)^2$, where the M_i are a series of different filters. The local energy model, in its use of two filters in quadrature, can be seen to be a specific case of quadratic filtering. Perona and Malik suggest that there is no special reason to use filters in quadrature and argue that one might wish to use quite different sets of filters. However, in the results they presented they chose to use two filters in quadrature; the second derivative of a Gaussian and its Hilbert transform.

Freeman, in his thesis [29] studied the local energy model with particular emphasis on multi-orientation analysis and the behaviour of local energy at feature junctions. He devised an approach to the detection and classification of feature junctions. The filters he used were generally second and fourth derivatives of Gaussians along with their corresponding Hilbert transforms, depending on the narrowness of the frequency tuning he required. As a tool for his multi-orientation analysis

Freeman developed the concept of steerable filters whereby filter outputs at any orientation can be efficiently computed from a linear combination of the outputs of a limited number of basis filters. Of relevance to the work presented in this thesis, Freeman developed a normalized measure of local energy. However, his motivation for doing this was primarily to allow image information to be represented over a small dynamic range rather than to specifically seek an invariant measure of feature significance. Some of his post-processing techniques might also be considered to be somewhat ad hoc. Despite this he considers a wide range of issues concerning the use of local energy for feature detection.

Ronse [78] makes a detailed mathematical study of the idempotency properties of the local energy model and the conditions of image modification over which local energy remains invariant. An important result, that will be used later, is that the locations of local energy peaks are invariant to smoothing of the image by a Gaussian or any other function having zero Fourier phase.

Rosenthaler et al. [79] make a comprehensive study of the behaviour of local energy at 2D image feature points. They develop a model of 2D feature detection based on differential geometry, using the first and second derivatives of oriented local energy to identify what they call *keypoints*. Robbins and Owens [76] have followed on from Rosenthaler et al.'s work and developed a simpler model of 2D feature detection that does not resort to the use of derivatives of the local energy signal. Instead, they detect 2D features by calculating oriented local energy over the image and then calculate local energy of this local energy image, but in an orientation perpendicular to the first. The second application of local energy detects the end points of any features detected by the first application of local energy. This process is then repeated over multiple orientations to capture all 2D features.

Wang and Jenkin [92] use complex Gabor filters to detect edges and bars in images. They recognize that step edges and bars have specific local phase properties which can be detected using filters in quadrature, however they do not connect the significance of high local energy with the concept of phase congruency.

One issue that previous work on local energy has not really addressed is the problem of how one should integrate data over many scales. If the perceptual significance of a peak in local energy is due to it also being a maximum in phase congruency then it is important to consider many scales simultaneously. After all, it is the occurrence of phase congruency over a range frequencies that makes it significant.

While the use of the local energy function to find peaks in phase congruency is computationally convenient it does not provide a dimensionless measure of feature significance as it is weighted by the sum of the Fourier component amplitudes, which have units lux. Thus, like derivative based feature detectors, local energy suffers from the problem that we are unable to specify in advance what level of response corresponds to a significant feature. Despite this, local energy remains a useful measure in that it responds to a wide range of feature types.

Phase congruency, on the other hand, *is* a dimensionless quantity. We obtain it by normalizing the local energy function; dividing energy by the sum of the Fourier amplitudes. Values of phase congruency vary from a maximum of 1, indicating a very significant feature, down to 0 indicating no significance. This property offers the promise of allowing one to specify universal feature thresholds, that is, we could set thresholds *before* an image is seen - truly automated feature detection.

2.4 Issues in calculating phase congruency

This section describes an initial attempt at devising a way of calculating phase congruency. What is highlighted is that there are a number of difficulties that have to be overcome if a practical method of calculating phase congruency is to be devised. These problems include the following: How should one extend the idea of phase congruency to 2D signals? What is the appropriate way of controlling the scale of analysis? How should information be integrated over many scales, and how can the influence of noise be overcome?

As mentioned earlier, phase congruency is awkward to calculate. An initial approach to calculating phase congruency might be to take a signal, remove its DC component, (it is removed because a 90 degree phase shift of a zero frequency does not have any meaning) calculate the Hilbert transform (say, by calculating the Fourier transform, multiplying the result by i and then performing an inverse Fourier transform), square and sum the Hilbert transform and the AC component of the signal, and finally normalize the result by dividing by the sum of the Fourier amplitudes. Results using this method were reported by Kovesi [47] (further work in which wavelets are used to calculate phase congruency were also presented by Kovesi [48]). An example of the calculation of phase congruency via the FFT is shown in Figure 7.

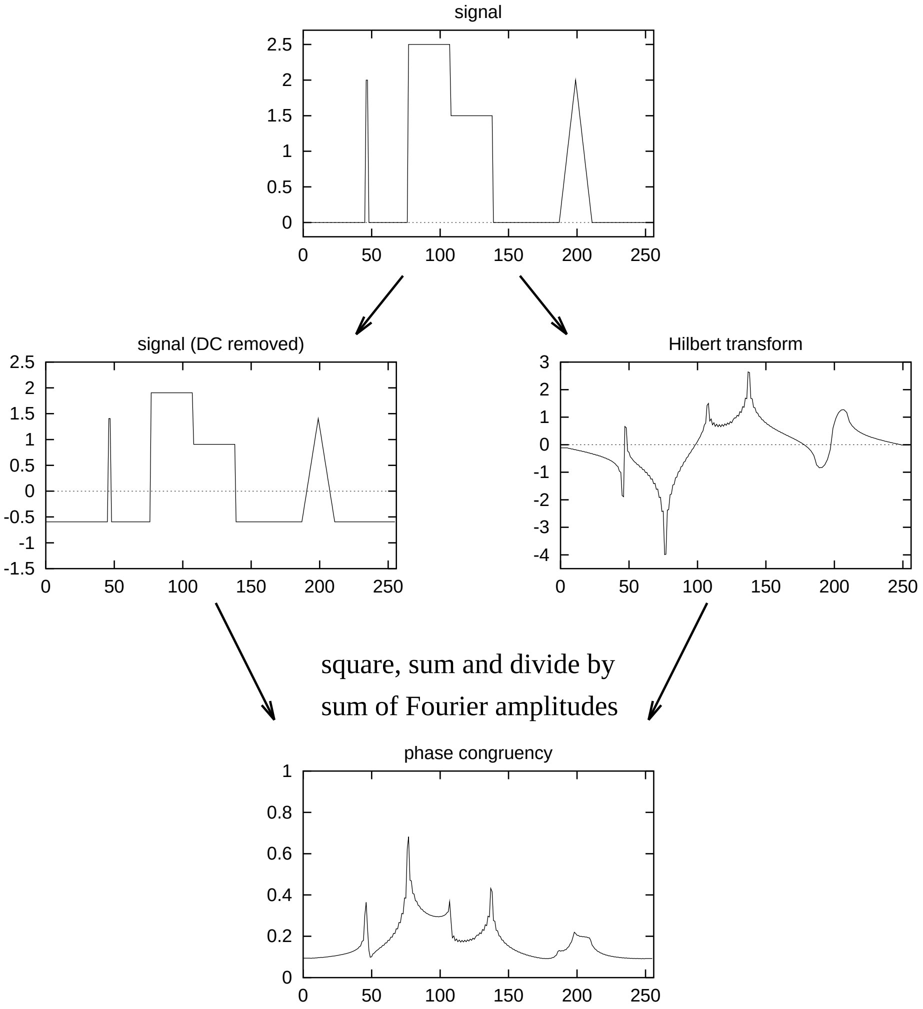


Figure 7: Calculation of phase congruency via the FFT. Notice how phase congruency values range between 0 and 1.

There are some problems with the calculation of phase congruency via the FFT. Firstly it is not clear how one adapts this approach for one-dimensional signals to two dimensions; the Hilbert transform is only defined in one dimension. A second difficulty is that the Fourier transform is not good for localizing frequency information spatially. In the example shown in Figure 7 the Fourier transform was calculated over the whole signal. Thus phase congruency at each point was calculated with respect to the whole signal. To control the local scale and spatial extent over which phase congruency is determined we have to use windowing of the signal. Windowing introduces the problem of having to balance spatial localization against the range of frequencies we wish to analyze; the window width controlling spatial localization but also constraining the lowest frequency we can measure. Figure 8 shows the result of calculating phase congruency using a rectangular windowing function 32 points wide. The computational procedure was as follows: Over each windowed section of the signal the Fourier transform was calculated, and the Hilbert transform generated. The signal value (minus the DC value) and the Hilbert transform value at the centre of the window was then squared and summed; this quantity would then be divided by the sum of the Fourier amplitudes over the current window to produce a phase congruency value at the centre position of the window. The window would then be indexed one point forward in the signal and the process repeated. Notice how the peaks in phase congruency are higher and more distinct. By windowing the signal each feature is considered in relative isolation to the others and hence ends up being considered to be very significant. An important point to note here is that for the calculation of phase congruency the natural scale parameter to vary is the size of the analysis window over which we calculate local frequency information. A large window means that the significance of features are determined in a more global manner, and a small window results in features being treated individually and locally. This leads to a new concept of multi-scale analysis which will be discussed in detail in the next chapter.

If the scale of analysis of phase congruency is controlled by window size we must consider what might happen when a windowed section of signal contains no features and only consists of noise. Being a normalized quantity, phase congruency does not depend on the magnitude of a feature on its own, it depends on the magnitude of the feature in the context of the local window. Thus, if the signal is purely noise each fluctuation in the signal will be considered quite significant relative to the surrounding features as they will all be of similar magnitude. Hence, noise poses

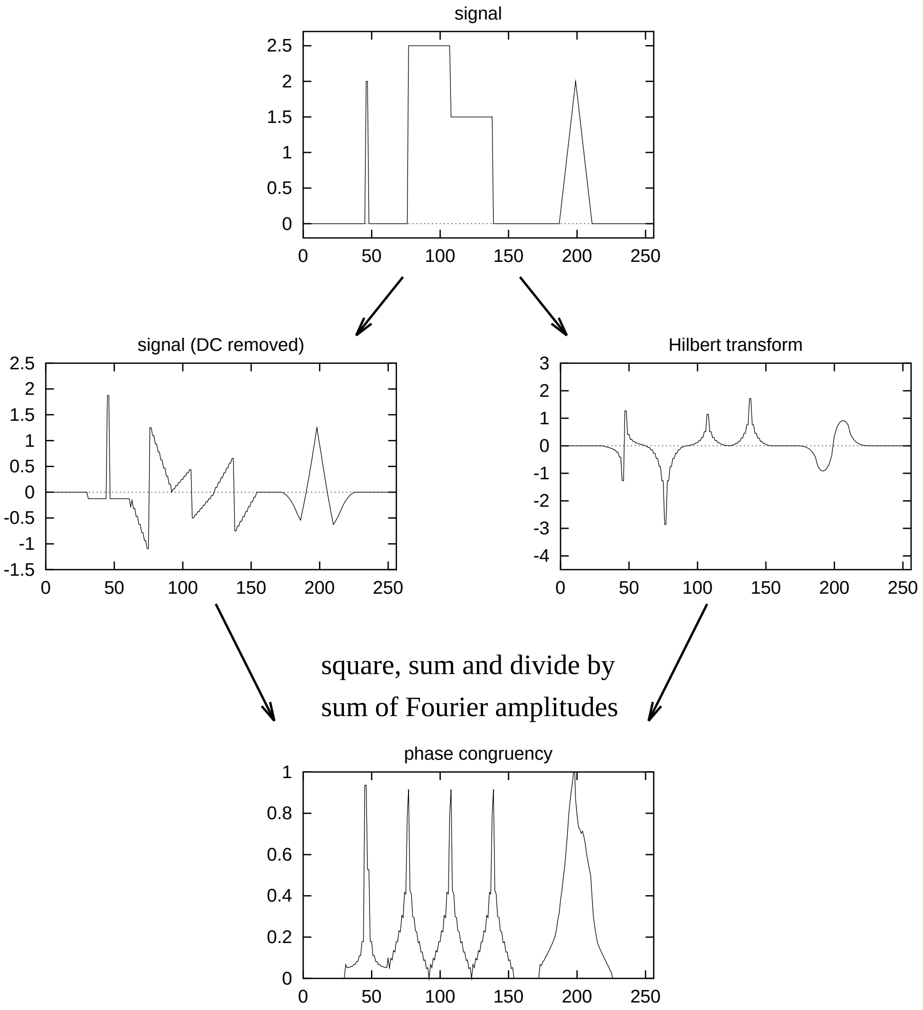


Figure 8: Calculation of phase congruency via the FFT using a rectangular windowing function 32 points wide.

a serious difficulty for us in trying to devise a practical way of calculating phase congruency in images. Figure 9 shows what happens if we introduce a small amount of noise into our signal. In regions that are distant from features the influence of noise becomes very noticeable.

A further issue we must also consider is that phase congruency as defined in Equation 3 does not take into account the *spread* of frequencies that are congruent at a point. For example, a signal containing only one frequency component, say a sine wave, will be in perfect congruence with itself and hence have phase congruency of 1 everywhere (the Hilbert transform of sine is cosine, and $sin^2(x) + cos^2(x)$ is identically 1 and so no point x has maximal local energy). To mark all such points as features would not make sense. Significant feature points are presumably ones