Content-based Image Compression

Michael Brady*, Paul Ducksbury†,
Vicky Mortimer*, Veit Schenk*,
Margaret Varga†, Bernard Liang*

1 Introduction

The rapidly increasing bandwidth of communications and the tumbling cost and size of electronic devices has led to an flood of data that shows no sign of abating. This has been compounded with increasing and increasingly severe demands from a wide range of applications for sensory information, particularly visual information. Examples abound:

- The huge investment in military sensors means that there is a rapid growth in the quality, diversity and quantity of images collected by military systems. This outpaces existing transmission, storage and retrieval systems. There is a military requirement for low bandwidth, covert operation, high quality transmission and large data volumes, all of which place severe demands on state-of-the-art technology. Consider, for example, the image shown in Figure 1. The military interest in this image is *not* the building in the centre of the image; rather it is in the detection and disposition of the aerials that surround it. Their detection is not at all trivial, as can be seen from the sub-window of the image centred on the rightmost aerial just above the building (see Figure 2). Conventional image compression inevitably blurs this out, as well as several other aerials, thus altering the apparent strategic import of the image. A second example makes a related, though different point. Consider Figure 5. An uninitiated observer is likely to miss potentially the most significant detail in the image: the engine bay of the sixth tank from the left is open, suggesting that it is in need of repair.

- Over the past five years, increasing numbers of hospitals throughout the world have installed PACS (Picture Archiving and Communication Systems) to transfer huge numbers of images directly to the workstation on the physician's desk. The growing number and variety of medical imaging systems mean that the clinician often has more image data to support diagnosis than he/she can handle in any reasonable time[1]. For example, x-ray mammograms are typically digitised to a resolution of 50 microns, so that each image is typically $4000 \times 4000 \times 2 = 32$ Megabytes. Since both cranio-caudal and mediolateral oblique views are routinely taken in screening centres in the UK, the image data gathered per patient per site is 128 Megabytes. Since there are typically 250 patients per screening day, each day generates about 32 gigabytes of data. Evidently, image compression is necessary. However, mammograms have relatively poor signal-to-noise ratio, suggesting that

*1. Department of Engineering Science, Oxford University; 2. DERA Malvern. Corresponding author: JMB

[1]This is the motivation for a recently announced inter-disciplinary research collaboration involving Oxford, King's, UCL, and Manchester universities.

the images be smoothed. Unfortunately, uniformly smoothing an image can change the appearance of a mass from that characteristic of benign to that of malignant. Conversely, mammograms have significant x-ray photon scatter [20], suggesting that the images be "sharpened". Unfortunately, this often has the opposite effect to smoothing: malignant masses can be made to appear benign.

- In response to rising levels of crime, the number of surveillance cameras has mushroomed over the past few years. The consequential flood of data is not entirely helpful because of the often poor spatial resolution of the images, the costs of storage and the equally important costs of retrieval. Note that the vast majority of surveillance data is of no use, since unwanted activities happen rarely.

- The internet and world-wide web encourages the transfer of images [52]. Uncompressed, the transmission of images over the internet, particularly as file attachments, takes an unacceptably long time. A number of software tools have been developed for file compression; but they are often ineffective on images, typically gaining a compression factor of only a few percent.

In all of the above cases, and in countless other application areas, image compression cannot be avoided in practice; but the information in the image that is important (for detection of malefactors, military decision making, and clinical diagnosis) is often subtle, not perceivable by the uninitiated, and is all too easily erased by conventional image compression schemes. In many applications, one cannot afford to apply a generic, lossy compression technique, as it may erase important detail. Consider, for example, Figure 3, in which the most information is the existence and position of the rail tracks. Unfortunately, as the reader may see by looking ahead to Figure 12, they are badly corrupted by most current compression techniques. Of course, compression techniques continue to improve [8; 19; 45], and we continue to make use of those improvements. (We have, for example, found that wavelet [48; 29; 9] and vector quantisation compression out-performs both JPEG and fractal compression on many images.) In this article, we report progress on a different, complementary approach to the problem.

We aim to compress an image by tesselating it into a set of disjoint subimages R_i and then applying a suitable, possibly different, compression technique $\mathcal{C}_i$ from a (possibly growing) set $\{\mathcal{C}_j\}$ to each individual region. Figure 4 illustrates the idea. Note that:

- The tesselation depends on the particular image, and is computed automatically using a combination of image analysis tools for region segmentation and knowledge-based techniques. In general, it is unlikely that the regions have simple geometric shapes (as is optimal for the vector quantisation compression technique). Rather, the regions correspond to objects in the image. The title of the paper reflects the fact that the way an image is compressed is determined automatically from the contents of that particular image. The word "contents" is deliberately ambiguous: it might equally refer to image analysis applied to that image, or it might refer to what the image connotes in the field of application.

- For the approach to be viable, we not only have to segment an image; but, given a particular subimage R, we need to be able to assess the effect of compressing R by each available compression technique $\mathcal{C}$. The assessment is not simply a compression ratio: a high compression ratio is of little comfort if the compression technique suppresses the information that is of most interest.

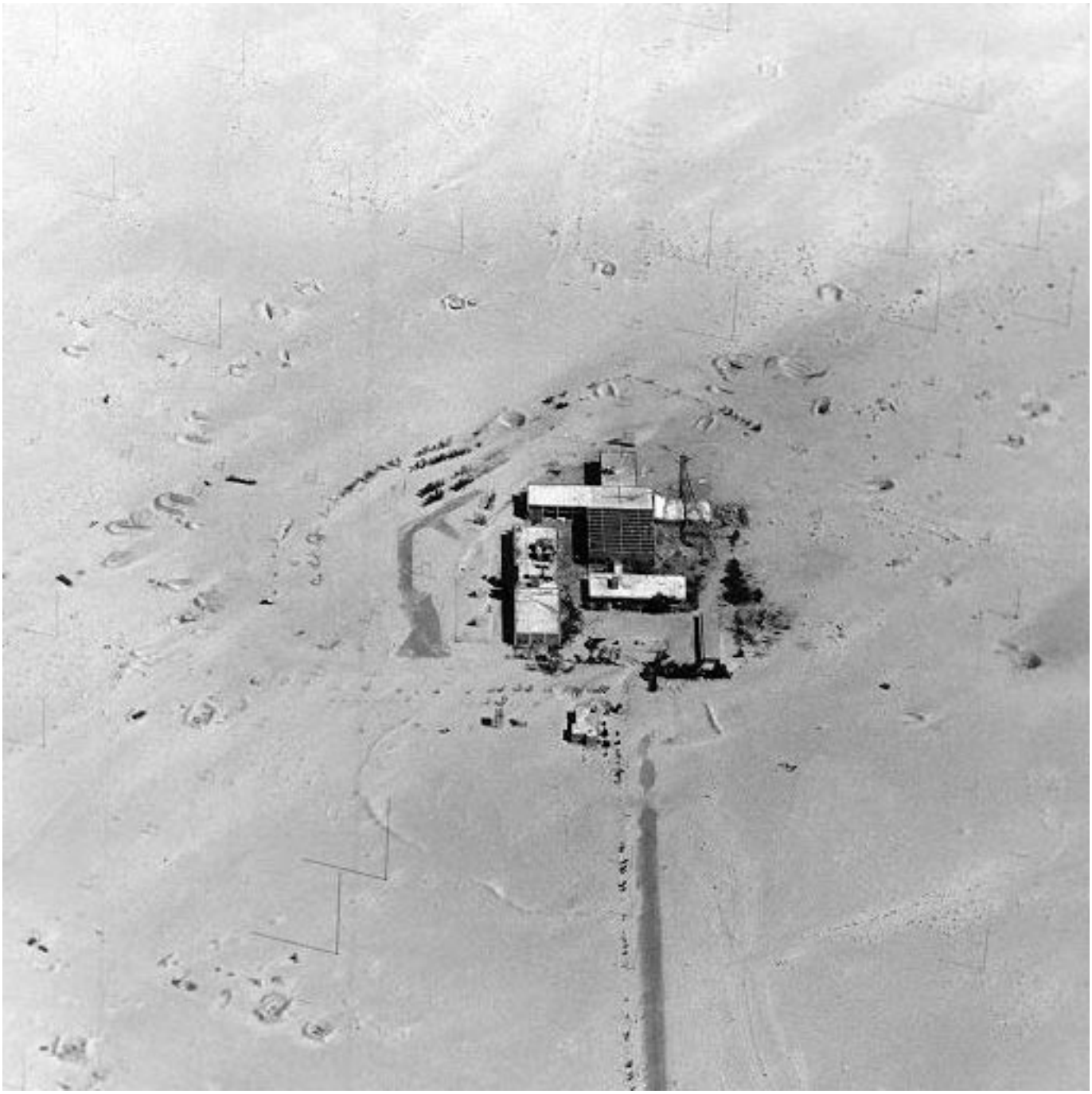


Figure 1: An image of a communications centre.

Figure 2: The portion of the image that we effortlessly perceive as the rightmost aerial just above the building.

Figure 3: In this image, the most significant information is the railway track. This is severely disrupted by most conventional compression algorithms, as can be seen in Figure 12

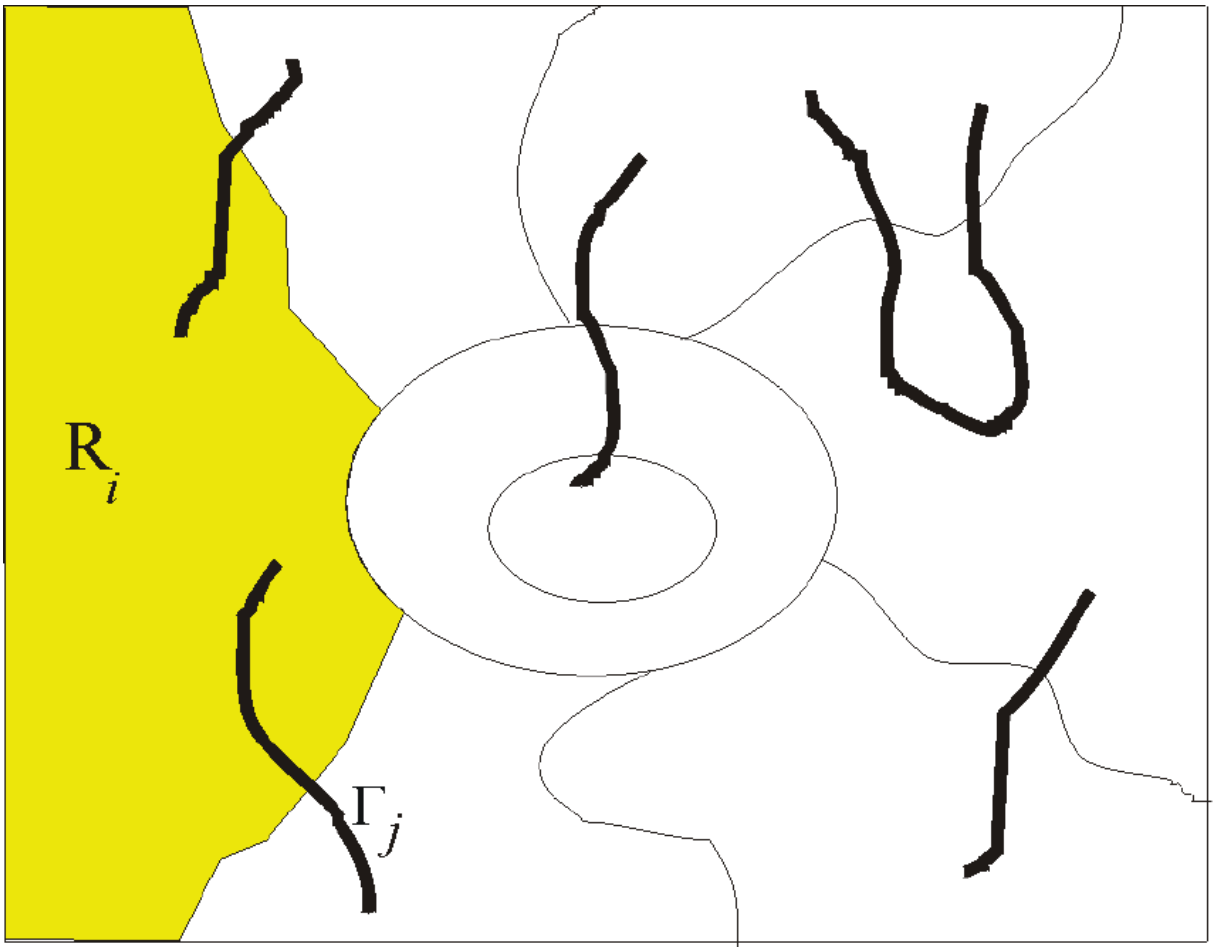


Figure 4: Cartoon of the tesselation of an image into a disjoint set of regions and open curves. The regions and curves are detected automatically according to the contents of the image. A coder is chosen automatically for each of the individual regions, while preserving the boundaries of the regions and the open curves.

- Some features that need to be transmitted without compression do not correspond to regions but to curves. These are typified by the aerials in Figure 1. Identifying such curve features is an important part of our approach.
- In the application areas with which we are most concerned (military and medical image analysis), there are often large areas that are of little interest and localised areas that are of great interest. By applying a suitable lossy compression technique to the areas of little interest and a suitable lossless compression technique to the areas of great interest, we aim to solve the problem of achieving significant compression ratios while preserving application-specific detail.

Figure 5: Aerial image of a set of tanks.

Our approach depends on high quality image segmentation into a set of regions that correspond to objects that are of interest in the particular application. This places our work at the intersection of image compression and image analysis, two fields whose interactions have, to date, been surprisingly few. To this end, we have developed a set of novel techniques for image segmentation using both "feature detection" and region segmentation. These two topics are the subjects of Section 2 and Section 3 respectively.

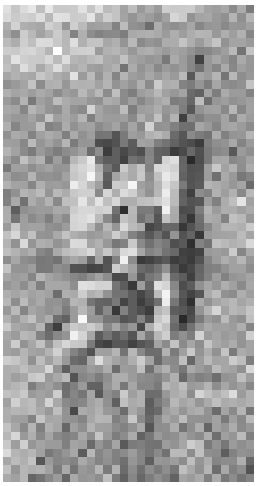


Figure 6: The most interesting aspect of this image is that the engine bay of sixth tank from the left of Figure 5 is open, suggesting that the entire fleet may be awaiting repairs to it.

Having tesselated the image into a set of subimages, the next problem is to choose a suitable compression technique for each. We introduce this problem in Section 5, in which we assess a range of currently available compression techniques on a typical image. Then, in Section 7 we show how entropy measurements can contribute to the choice of compression technique for each subimage. We conclude in Section 8 with a discussion of the potential for applying artificial intelligence techniques to add application-specific knowledge to the entropy-based analysis.

2 Feature detection

Gradient based Edge Detectors

Computer vision has developed a set of image "feature" detectors that detect and encode intensity changes. This is important because, as has frequently been observed, human visual perception depends critically upon the detection of intensity changes. We effortlessly glean a remarkable amount of information from a line drawing sketch of the intensity changes in an image, suggesting that a line drawing might be a massively compressed version of an image that nevertheless preserves all of the important information. This idea is further reinforced for some scientists/engineers by the observation that the human visual system dedicates enormous resource apparently towards producing exactly such a sketch [30]. The success of subsequent image analysis processes, such as matching, recognition, and motion estimation, is often determined to a great extent by the number and quality of the features detected. The same consideration applies equally to image compression.

Until recently, feature detection has largely comprised "edge" detection [41; 7; 10] and "corner" detection [18; 55; 49]. Edges are usually considered to be intensity changes that are locally one-dimensional and whose profile in the orthogonal direction approximates a step change. Corners are intensity changes that are locally two-dimensional. Most effort to date has been aimed at edge detection.

More precisely, the designs of most "edge" detectors have essentially been optimised for the detection and localization of step changes in intensity. In essence, this amounts to a variety of proposals for estimating the amplitude of the directional derivative of the intensity function at

each image point, after suitable (e.g. anisotropic) smoothing to regularise the computation [34].

Composite Features

Unfortunately, many significant intensity changes in the applications of interest are not at all step-like. Consider, for example, the intensity change corresponding to the aerial shown in Figure 2. Similarly, many of the intensity changes of most interest in Figure 3, particularly the railway tracks, are not step-like. Indeed, idealizations of intensity changes that occur regularly include: steps; "roof" changes comprising a forward ramp followed by a downward intensity ramp; "thin bars" in which two back-to-back steps are spatially close; and textural variations. Typically, the responses of feature detectors optimised for steps often degrade hopelessly when applied to non-steps. One early approach was to develop a set of feature detectors, each optimised for a different idealised intensity change, and then attempt to combine their outputs in order to construct a composite representation of the intensity changes of all types. The most comprehensive proposal of this sort was the *Primal Sketch* by David Marr [30]. Unfortunately, real intensity changes are often complex combinations of the idealised changes referred to above, and (i) the individual detectors often give poor response to such composite changes, and (ii) the rules for combining feature detectors quickly become very complex. There has been little progress with this approach since Marr's pioneering work.

Cross-section independent Feature Detection: Local Energy and Phase

In essence, the above approach is rooted in mathematics: the aim is to match the image-signal to a function, the closer the signal to the function, the higher the response.

An alternative approach is to take as a starting point what the human visual system HVS might consider "interesting" as opposed to what "matches a certain function". It turns out that the HVS responds strongly to high intensity gradients, which is why gradient-based edge detectors are so popular. One theory is that the HVS responds to any part of a signal which has high local energy.

Local energy is defined as follows:

$$LE(x) = \sqrt{O(x)^2 + E(x)^2}, \tag{1}$$

where O and E are odd and even-symmetric filters in quadrature, i.e. they are phase shifted by $\frac{\pi}{2}$. Two such filters will always respond to both the odd and even-symmetric parts of a signal and hence give a response which is independent of the actual shape of the underlying signal.

A theory of feature detection which has emerged over the past thirteen years is that there appears to be a single characteristic that *all* image features (luminance profiles that humans perceive as places of interest) have in common, namely that the frequency components of each of these signals have approximately the same phase-value over a wide range of the frequency-spectrum [36]. In other words, the phase-values are "congruent" over a range of frequencies.

This is made precise below. Equally importantly, the actual angle at which this phase-congruency occurs is characteristic of the type of feature. Thus for a step up in intensity, all the local (windowed) Fourier components have phase zero, for a step down they all have phase π, for an up-pointing intensity ridge (a roof) they have phase $\frac{\pi}{2}$, and a down-pointing ridge (the so-called valley detector used to good effect in telephone coding schemes for sign language [38; 40; 39]) has phase $\frac{3\pi}{2}$. This observation suggests that the imprecise perceptual concept of a

"feature" may be precisely *defined* as an image location at which there is a local congruence of phase.

Reinforcing this idea is the fact that it is well known that phase carries (most of) the important information about a signal or image [35; 6]. It is in fact quite astonishing to observe that the overwhelming majority of feature detectors developed to date for image analysis retain only amplitude information. As Kovesi [24] has pointed out, this not only causes them to fail to detect features of interest, it also reduces markedly their invariance to intensity contrast, and makes them sensitive to thresholds.

An Implementation of Phase-congruency: PC

More precisely, let $I(x)$ be a one-dimensional signal (we return to the case of two-dimensional images below). Suppose that the short-term Fourier Transform (STFT) expansion of $I(x)$ at location x is given by:

$$I(x) = \sum_{n=1} A_n \cos(n\omega x + \phi_{n_0}) = \sum_{n=1} A_n \cos(\phi_n(x)) \tag{2}$$

where $A_n, \phi_n(x)$ are respectively the n^{th} components of amplitude and phase. Kovesi proposes a measure $PC(x)$ of phase congruency at signal location x by:

$$PC(x) = \max_{\overline{\phi}(x)} \left[\frac{\sum_{n=1} A_n \cos(\phi_n(x) - \overline{\phi}(x))}{\sum_{n=1} A_n} \right] = \frac{\mathcal{E}}{\sum_{n=1} A_n} = \frac{Num}{Den} \tag{3}$$

PC is a dimensionless quantity with a value between 0 and 1. It can be shown that the numerator E in the above expression is the local energy of the signal [54]; PC is the (local) maximum of the amplitude-weighted sum of the phases (computed over a range of scales) normalised by the amplitude sum. In theory, PC is invariant to local image contrast. That is, PC will generate the same response to features irrespective of the contrast, magnification and brightness of the image, making the compression coding scheme robust against thresholds. In the following, it is important to distinguish between the *concept* of phase-congruency and Kovesi's *implementation* of PC as defined above.

PC vs Local Energy

Local energy and phase-congruency are in fact directly linked. This was demonstrated by Owens and Venkatesh [54]: phase-congruency is proportional to local energy. However, this relationship only holds in one direction: high local energy always implies high phase-congruency. However, the reverse is not always true: a signal can have high phase-congruency, but low local energy. Examples include signals with a very low frequency spread or low-amplitude signals embedded in noise: the very fact that humans turn to computers to "improve" fuzzy, low contrast images shows that the HVS is quite poor at detecting signals with a low SNR, despite the fact that the features contained in the signal may have high phase-congruency. In other words: phase-congruency is necessary but not always sufficient.

Another confusing fact is that when the signal is filtered with a bank of bandpass filters, each individual response is directly related to a measure of phase-congruency in itself. This is simply because each filter computes a local-energy measure, which, as pointed out above, is directly

related to phase-congruency. PC then is a normalised weighted average of weighted average measures of energy contained in individual frequency bins. As a consequence, we contend that for feature detection, the local energy approach may be the more appropriate method of the two.

In fact, we observe a number of difficulties that occur in practice in the implementation of PC, even when it is to be applied to one-dimensional signals.

A normalised Measure?

The quadrature filters O and E mentioned above are $\frac{\pi}{2}$ out of phase, i.e. they are a Hilbert transform pair. Since the Hilbert transform is not defined for two-dimensional signals, images are typically convolved with pairs of filters which are (approximately) $\frac{\pi}{2}$ out of phase. As a result, their responses depends strongly on the local neighbourhood in which the measure is computed: adjacent dominant features influence, and may even suppress, the response to minor features. As a result, the response of PC is always be dependent on the underlying signal and can never be independent of contrast, brightness etc, which it claims to be.

Thresholds

PC requires that a number of thresholds be set, principally the choice of range of scales to be used to compute the PC measure. This range of scales depends critically on the underlying signal. Fully automated feature detection system that operates without human intervention do not yet exist. As an example, for a signal containing predominantly large-scale features one would select a different range of scales than for a signal containing many important small scale features. In particular the first application would have a significantly "larger" minimum scale than the second. This has important consequences for noise suppression.

Noise

Precisely because PC aims to be invariant to local image contrast by normalising the local energy measure, it is as sensitive to noise as it is to signal. Noise estimation and suppression are key to making PC useful in practice. In Kovesi's implementation of PC, for example, the amplification of noise is addressed by statistical noise-suppression based on the output of the first level of the filter-bank used to build the multi-scale representation of the signal. Different features or feature types occur at different levels of the multi-scale representation, so we should not treat the signal uniformly using only the first level of the representation. Another question is how to choose the first level. What if there are only large-scale features in the signal? In that case we should use only relatively low-frequency filters. But this means that we cannot use the first level (which in another scenario would be the n^{th} level) for noise-estimation. In short, we need to treat image-noise, filter noise (independent of filter level) and possibly image-noise due to the sampling grid for orientated filters, but not propagate conclusions drawn from one level to all others.

Choice of Scales

Imagine now an application in which the signal contains a mix of large and small scale features, some of which are small scale features superimposed onto large-scale features (i.e. they are composite features). All such features have their local energies distributed and weighted

differently. As an example, a relatively wide bar will have a significant local energy contribution from the filter that is approximately of the same size as the bar. But there will also be energy contributions from filters which respond to the two edges of the bar. A thin bar (or line), on the other hand, will only have a strong energy contribution from the small scale filter that responds to the thin line.

This indicates that simple averaging (amplitude weighted or not) of filter responses over the *whole* range of frequencies, as effected in Kovesi's implementation of PC, might not be appropriate to extract general features reliably from images. Figure ?? illustrates this fact: when using predominantly low frequencies (shown in magenta), the large (left) negative bar of the incision is marked correctly whereas when using mostly high frequencies (red), the three step edges of the incision are marked. The average PC taken over all scales is shown in blue: clearly this is *not* the desired response (although in practice, sufficiently isolated edges are detected reasonably well).

We address this problem using an approach similar to Lindeberg's scale selection mechanism [27; 28]: for composite features, only certain ranges of scale have significant local energy. For a highlight adjacent to a shadow, as shown in figure 7, there is significant local energy at small spatial scales for each of the individual steps but also varying amounts at larger spatial scales when these individual features merge. Yet PC, as defined above, only gives significant responses to the individual step edges and only very weak responses to the underlying structure, which is two adjacent bars (one negative, one positive). By automatically selecting:

1. the range of scales over which there is a significant response (local energy)
2. the point within this range which corresponds to the local maximum, and
3. the 'width' (or 'blobsize' using Lindeberg's terminology) of such a response in the multi-scale representation,

we obtain information concerning not only the size, extent and the degree of smoothing; but also to what extent a feature is part of a larger composite feature. Using this type of automated scale detection and selection also address the two shortcomings of PC mentioned above: (i) the setting of thresholds, or in other words selecting the 'right' range of scales in advance; and (ii) the problem of identifying and isolating spatially close features which may or may not form part of a composite feature.

Orientation

Kovesi's algorithm is based on quadrature log-Gabor filters, so it intrinsically makes appeal to the Hilbert Transform [4]. It is well-known that there is no equivalent of the Hilbert Transform in two or more dimensions, and so the 1D PC algorithm (signals) presented above does not have a straightforward extension to 2D (images).

The computation of $PC(x)$ for a two-dimensional image requires the combination of a number of values $PC_i(x)$ computed by the one-dimensional algorithm in a suitable number of directions i at each image point x. The optimal way to do this remains an open problem, to which we return below. Kovesi suggests that 6 directions is a good compromise between uniform response and excessive computation. He computed the average $NumAvg$ over all directions i of the numerators Num_i, and the average $DenAvg$ of the denominators Den_i. Finally, his composite PC measure at an image point x was $\frac{NumAvg}{DenAvg}$.

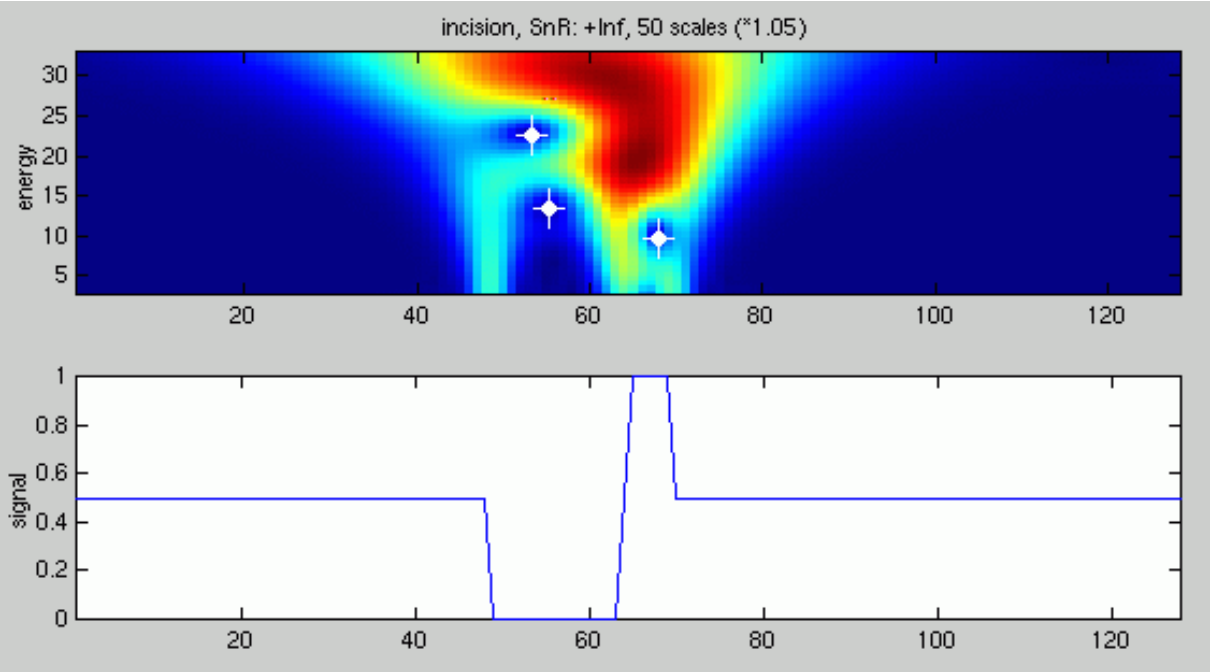


Figure 7: Singularities (marked) occur where two features merge

There are two problems with this measure. First, the ratios of the individual Num_i and Den_i can be small even if the individual values are high, suggestive of a feature. Second, we expect to see differences between edges and corners. In essence, the distributions of the numerators and denominators are often highly correlated. While there may be good results for summing (in effect averaging) the oriented 1D PC results over orientations, for example to retain the analogy of "energy" and to compensate for the uneven response over directions, averaging eliminates relatively faint features. More importantly, averaging fails to exploit the characteristics of features encoded in $\{Num_1, Num_2..Num_6\}, \{Den_1, Den_2..Den_6\}$. As Gilles [16; **?**] notes, an edge is locally "simple" whereas a corner/junction is "complicated", an intuition that can be made precise, as is discussed in [15]. Edge points might be expected *a priori* to give rise to a peaked unimodal distribution, whereas corner points would be expected to have a distribution that is less peaked but still show substantial contrast with orientation. Finally, background points should in general have a flatter distribution. These observations motivated an investigation of alternative techniques for combining the directional estimates of PC [26], leading finally to a rule-based system for distinguishing a range of feature types. Replacing the average numerator divided by the average denominator as the 2D measure of PC by non-linear combinations that seek, for example, directional maxima with a minimum in the orthogonal direction, enables many of the aerials in Figure 1 to be detected: see Figure 8.

Local Energy and Coding

The massive and unique advantage of the local energy (or, to a certain extent, phase congruency) approach to feature detection is that it responds to a wide range of feature types. This can be exploited in a coding/compression setting. In principle, one computes the local

Figure 8: The feature points detected by a nonlinear combination of directional phase congruency responses. See text for more details.

energy (or PC)-map for an image, then keeps those areas that are considered important, the rest could be discarded (coding: produce line-drawings) or compressed. Many features will bound regions (q.v.). Others correspond to thin directional features (in aerial images: pipelines, telegraph/power-lines, railways etc; in medical images: blood vessels, ducts, stroma) which are (depending on scale) essentially delta-lines. These thin lines can be preserved (either in a line drawing or by not compressing them) whereas other areas, such as they sky/sand etc could be removed/compressed.

3 Image segmentation

The goal of image segmentation is to tesselate an image in to a set of compact, simply-connected subsets ("regions") R_i, each of which is homogeneous in some suitable image property such as intensity, colour, texture, or motion [2; 37], and such that the homogeneities of each pair of contiguous regions R_i, R_j are sufficiently different. The boundary of each region R_i is a closed curve. Local variations in pixel properties, due both to imaging noise and to the textures of visible surfaces, mean that in practice, "homogeneity" does not equate to a single value; but to an appropriate probability density function (pdf). It then follows that "sufficiently different" equates to (statistically) significantly different.

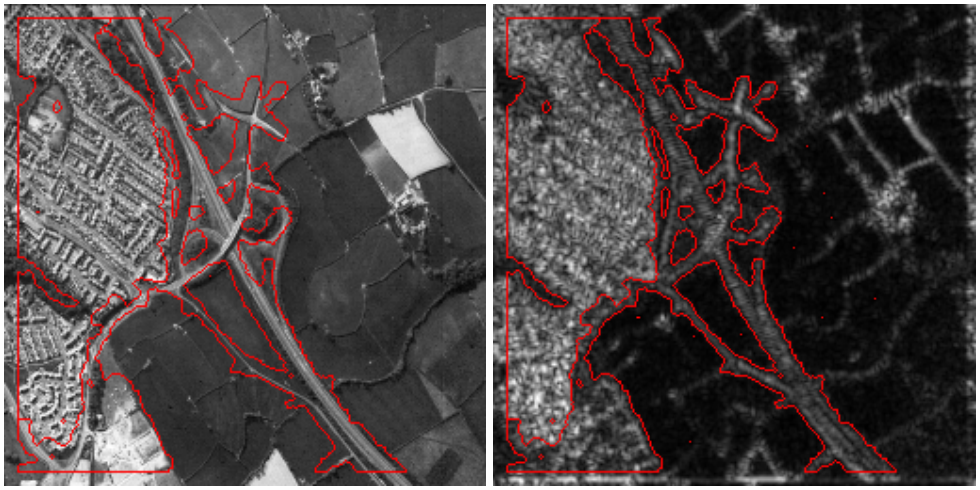
In the special case of images that are piecewise constant in intensity (or colour), region boundaries correspond to step changes in intensity (or colour). For this reason, it is often asserted that image segmentation is the flip-side of feature detection. Indeed, Marr [30] considered image segmentation to be mathematically ill-defined and did not include it in his theory of vision. The previous section should have convinced the reader that there is considerably more to feature detection than step edges. In practice, the boundaries between two textured regions are often difficult to detect using even the most advanced feature detectors. In particular, it is often the case that feature detectors generate portions of contours that approximate well parts of region boundaries; but do not in general yield closed contours, which regions are required to have.

With this in mind, approaches to image segmentation can be roughly classified into three groups:

- Active contour methods [3], such as snakes [23] and balloons [60];
- Region split-and-merge techniques [1];
- Global optimisation approaches based on energy functionals [33], Bayesian analysis [14] and/or the Minimum Description Length (MDL) technique [22].

Each of these approaches has intrinsic limitations. For example, snake or balloon models only make use of information along region boundaries, so that relatively little (or, in the typical case, no) information from the interior of the region is considered. An advantage of region growing is that it tests the statistics inside the region, however it often generates irregular boundaries and produces small holes. On the other hand, energy/Bayes/MDL utilise global criteria; but it is often very difficult to escape local minima in the associated search process.

Region competition was developed recently [61] for image segmentation. It combines a number of the attractive geometric features of the snake/balloon approach with the statistical techniques of region growing in order to locate the "best" positions for the boundary between neighbouring regions. It builds upon Leclerc's [25] influential idea of minimum description length (MDL) coding of an image. The basic idea is to iterate from an initial tesselation of the image,



(1) *(2)*

Figure 9: *An example of applying region segmentation. (1) Results overimposed in the original texture image; (2) Results overimposed in the wavelet local energy image..*

computed, for example from a suitable set of seed points, by a region growing and merging algorithm. The regions in the tesselation then compete for points on the boundary of the tesselation in an algorithm that attempts to compute a MDL code for the tesselated image. This is an optimisation algorithm that minimises three kinds of cost:

- a fixed cost for each region: this encourages tesselations with as few regions as possible;
- a cost for the contour of each region, which discourages substantial curvature variations along the contour; and
- a "statistical force" cost that uses a suitable statistical model of the points inside the (current) region to measure the likelihood that a given boundary point belongs on one side of the boundary rather than the other (see [61; 5] for further details).

In previous work, we have extended region competition algorithm in two ways in order to achieve good results on textured (primarily linescan, infrared aerial) images [5]. First, we have developed a novel texture descriptor by computing wavelet local energy, at a set of spatial scales, in the horizontal, vertical, and diagonal channels of the wavelet multiresolution decomposition of the image [29]. Second, having noted that the distribution of the local energy descriptor in perceptually homogeneous textured regions often does not have a distribution that conforms to a parametric statistical distribution, we have used the Wilcoxon-Mann-Whitney U statistic in the "statistical force" of the region competition algorithm.

The combined algorithm has been applied to a large number of textured aerial images. Figure 9 shows a typical typical aerial image and the corresponding segmentation result (several other results are shown in [5]).

4 Filter fusion

Schemes for feature detection and image segmentation continue to improve, as illustrated by the two previous sections. However, no image analysis scheme that has been developed to

date works uniformly well on all images, even on all images from a single application. We have further developed the texture segmentation algorithm sketched in the previous section, the feature filtering approach described in Section 2, and on a number of other segmentation techniques that we had developed previously, by attempting to combine, or "fuse", their results.

In particular, we have explored in a variety of applications the use of Belief Networks for combining evidence [11; 53; 12]. A description of belief networks, in particular their application to texture segmentation, is beyond the scope of this paper and can be found in the cited references. The idea is to combine the evidence provided by the individual "filters" into a belief of an appropriate hypothesis. Typical hypotheses might be that a given region of an aerial image corresponds to an urban area, or, in a mammogram, corresponds to a malignant mass.

The evidence to be combined typically ranges from elementary (and fast to compute) statistical/texture measures through to more advanced higher level evidence generated by specific feature detection processes. The belief network approach can be used as a feature detection process in is own right or as a cueing mechanism for other algorithms such as region competition, as described above.

Figure 10: Example of the use of multi-level belief networks to fuse information from a set of feature detectors.

Figure 10 shows an example in which a multi-level belief network has been used to fuse information from a set of simple feature detectors in this case edges, extrema, and a classification of the gray level distribution using a chi-squared measure. Three possible states (low, medium, high) for the hypothesis 'urban region' are used with an outline being drawn around the 'high'

state. This type of approach has been used as a feature detector in its own right as well as a cueing mechanism for other more specific algorithms.

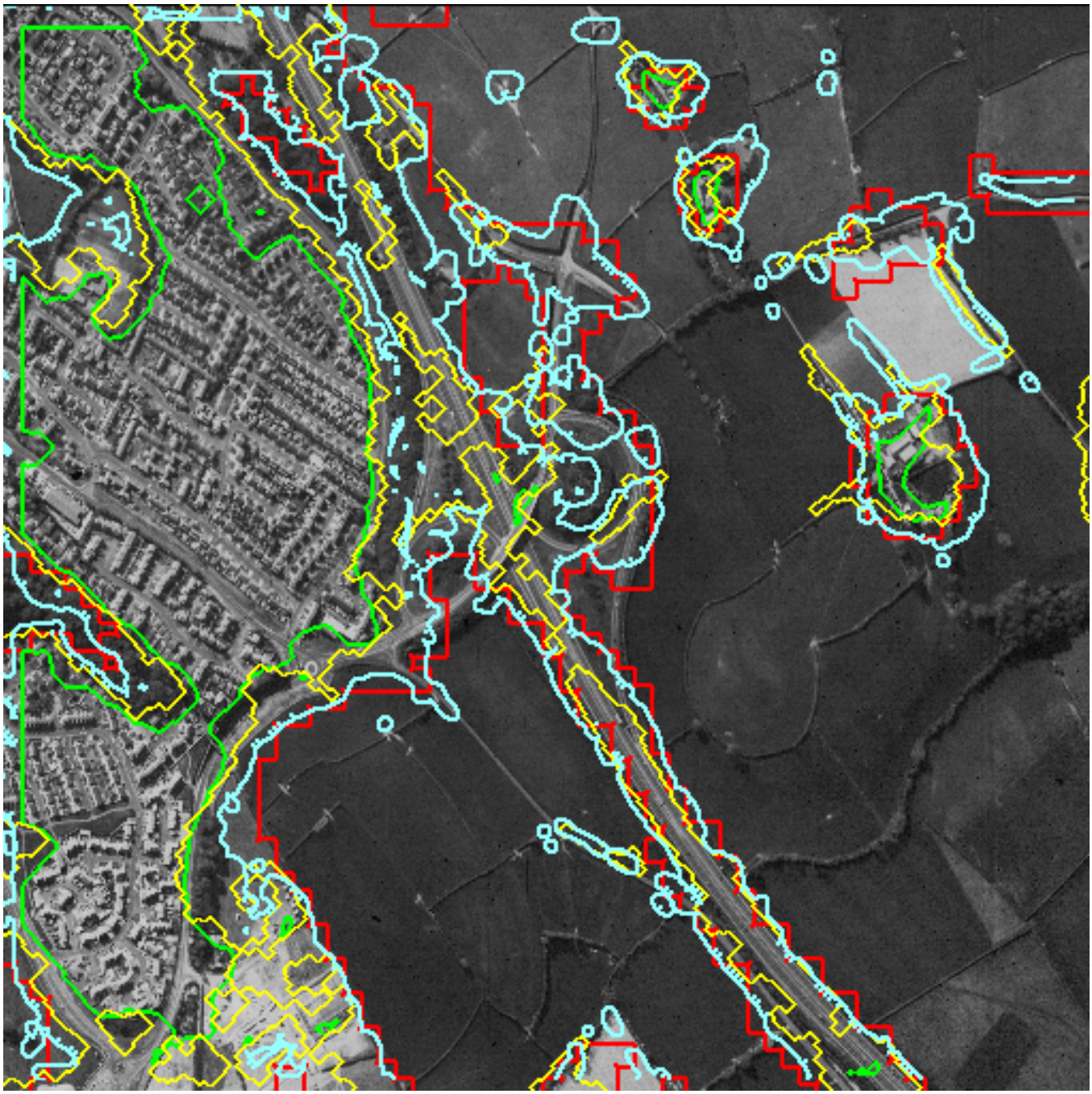


Figure 11: Probabilistic feature detector fusion. Three different segmentations, shown in red, green and yellow were fused using a belief network algorithm.

Figure 11 shows a second example in which three the results of applying three different region segmentation algorithms were fused to give a composite segmentation. The segmentation shown in red was generated using the same belief network used in Figure 10 [53]. The segmentation shown in green resulted from the application of a fractal analysis of the image [59]. The

segmentation shown in yellow resulted from a wavelet local energy model of texture [58]. The three resulting probability images were then fused using a second belief network [53], resulting in the segmentation shown in pale blue. Note that some pale blue regions are present where there appears to be no input evidence! This is, in fact, not the case; the three coloured contours (red, green, yellow) correspond to binarised visualisations of the three segmentations and are each created from probability surfaces. In each case, evidence exists for the fused segmentation and has been combined appropriately.

5 Compression techniques

The previous three sections report progress on the first part of our approach in that they enable (i) certain disconnected curves (typified by the aerials) to be extracted and coded explicitly and (ii) enable textured images to be segmented into regions of uniform texture. The aim is then to code the regions and preserve the curves that have been detected. Our strategy for coding will be described in the following section. First, however, we pause to show typical results of applying current coding techniques to images of interest for the current application.

Figure 12 shows the same image after compression and subsequent reconstruction using four of the currently best known image compression techniques (one technique per row) at each of three compression ratios (columns). The compression techniques, in order from top to bottom, are: JPEG, fractal, wavelet, and vector quantisation. The three compression ratios are, from left to right, 10:1, 30:1, and 50:1. The image is of interest because it contains: textured regions, the featureless "apron" region that the tanks are standing on, objects with strong features (the tank boundaries), and features that do not correspond to step changes in intensity (the railway lines) and which need to be preserved by the coder/decoder.

The top row of Figure 12 shows the results of applying JPEG coding. The results are quite typical of JPEG: even at low compression ratios, it develops a characteristic "blocky" appearance. Note also that, again at quite low compression ratios, it suppresses important edge information - in this case the railway lines. The second row of Figure 12 shows the results of applying fractal compression. This technique is primarily intended for use with textured regions, many of which can be well approximated by fractal measures [59]. Unsurprisingly, it gives poor results on regions that are not well approximated as fractal surfaces, for example the apron region and the tanks. Generally, fractal coding performs equally poorly on regions that are sparsely populated with features, such as the railway track. Here it performs no better than JPEG. The third row of Figure 12 shows the results of applying wavelet coding [19]. It can be seen that the result is generally acceptable up to 30:1 compression (the second column); but that the railway line and surrounding texture breaks up by 50:1 compression. Finally, the bottom row of Figure 12 shows the results of applying vector quantisation (VQ) [8]. On this image, it gives the best results of the four approaches studied. The edges of the tanks are preserved quite well, as is the railway line, though the latter has begun to break up by 50:1 coding. In general, VQ and wavelet coding is acceptable for this image, certainly up to 30:1 compression ratios.

Closer inspection suggests that, at 30:1 compression, the VQ codec results are slightly superior, having less artifacts such as ghosting around structures. We find that both the VQ and wavelet approaches can achieve 100:1 compression, in the sense that objects of interest are still interpretable by the human visual system. However, we can already conclude that there are systematic failures of both coding schemes near perceptual features and that this can suppress information that is important to human analysis of image content.

To gain a little more insight into the differing performances of the algorithms, in particular the wavelet and VQ coders, consider Figure 13. The columns are different compression ratios, as per Figure 12. The rows are the following:

1. *raw image vs wavelet coding*: The top row of Figure 13 shows the difference image between the raw image (which corresponds to "ground truth" insofar as it represents the case of no coding at all) and the result of wavelet coding the image and then decoding it with the same wavelet. More precisely, the first row of Figure 13 shows

$$I_{\mathrm{raw}}(x,y) - decode_{\mathrm{wavelet}}(code_{\mathrm{wavelet}}(I_{\mathrm{raw}}(x,y))).$$

   Notwithstanding the comments above, which reflect our human perception of the appearance of the coded/decoded images, the difference image is a good representation of where information is suppressed. If the codec only suppressed local image variations (for example, noise), then there would be no structural information perceivable in the difference image. That is not the case. The thin lines corresponding the railway are clearly visible at each compression level, and the outlines of the tanks are clearly visible even at compression level 30:1.

2. *raw image vs VQ*: the result of applying the same process to VQ coding is shown in the second row of the Figure 13.

   Substantially the same comments apply as in the case of the wavelet codec. Again, the structure that corresponds to the railway is clearly visible, as is the upper left tank. On the other hand, there appears to be less structure in the difference image in the vicinity of the other tanks than for the wavelet codec. This demonstrates how we can be deceived by the exquisite capabilities of human vision if we only analyse visually the results of applying a coding scheme.

3. *VQ vs wavelet*: the third row of Figure 13 shows the result of computing

$$decode_{\mathrm{VQ}}(code_{\mathrm{VQ}}(I_{\mathrm{raw}})) - decode_{\mathrm{wavelet}}(code_{\mathrm{wavelet}}(I_{\mathrm{raw}}))$$

   The bottom row of Figure 13 compares the results of applying the wavelet and VQ coding algorithms. The interesting aspect of this difference image is that even at a casual glance it contains perceptual structure, unsurprisingly corresponding to the railway.

The conclusions from this experiment, and from the many similar ones that we have conducted on a range of imagery, are that:

- conventional coders suppress structural information in images that is often key to the subsequent use of that image;
- visual comparison of the original to the results of applying a codec cannot be wholly relied upon; and
- different codecs perform more or less well on different regions of an image.

The next section oulines the challenge and the following one outlines the second part of our approach by addressing the latter point.

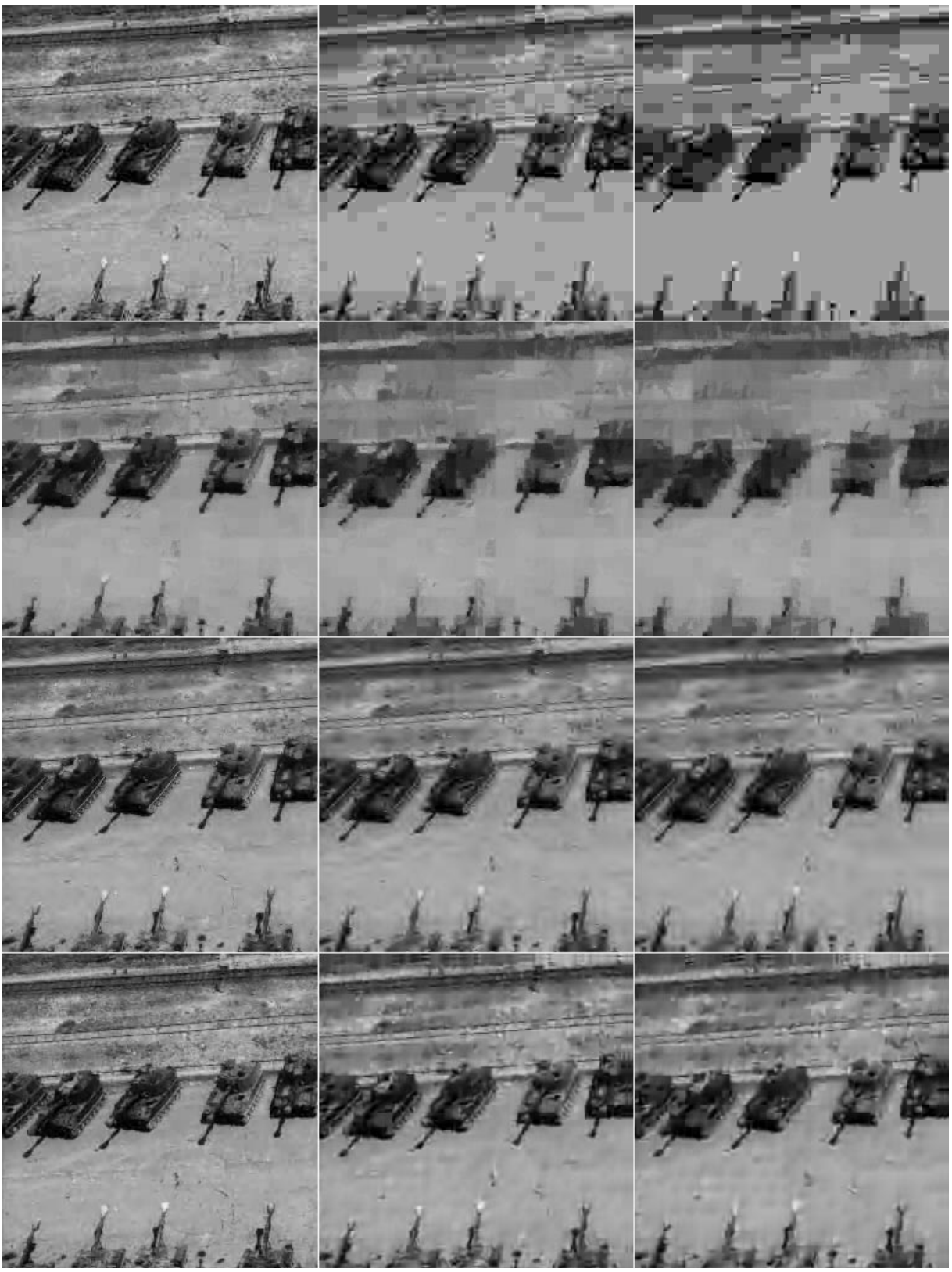


Figure 12: The same image is shown after compression and subsequent reconstruction using four of the best known compression techniques (rows) at each of three compression ratios (columns). The compression techniques, in order from top to bottom, are: JPEG, fractal, wavelet, and vector quantisation. The three compression ratios are, from left to right, 10:1, 30:1, and 50:1.

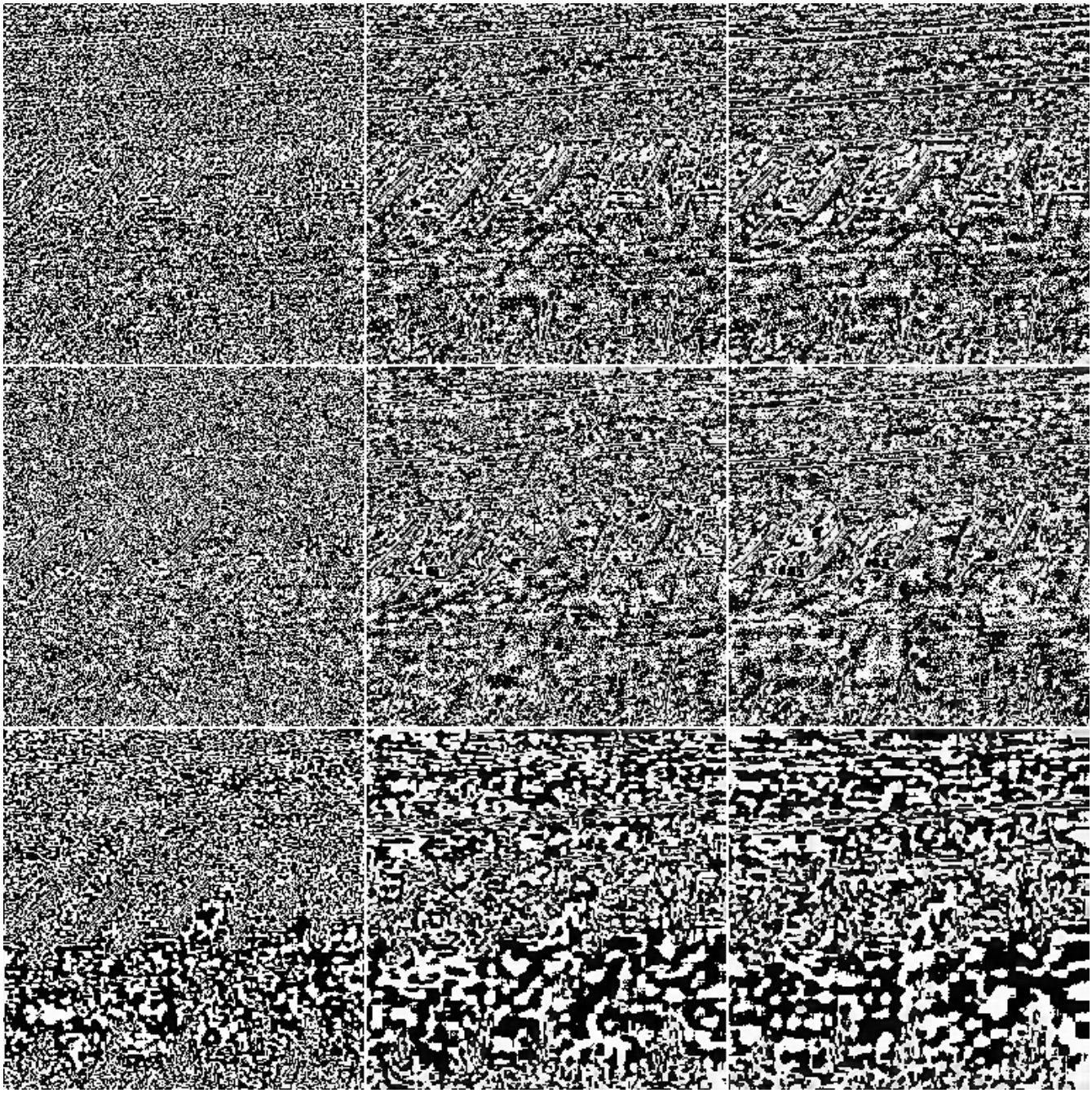


Figure 13: The results of comparing code-then-decode (codec) for wavelet coding is shown in the top row for the image shown in 12. The second row is the same process applied to VQ, and the third row is the difference between VQ and wavelet codecs. See text for detail. The three compression ratios are, from left to right, 10:1, 30:1, and 50:1.

6 Performance Evaluation

In practice, the success of an image compression scheme may be measured in terms of the average or worst case compression ratio it can achieve on the images of interest, on its speed, and on the resulting image "quality". Whereas compression ratio and speed can be measured accurately, image "quality" is considerably less straightforward, not least because the perception of "quality" will vary according to the expertise and purpose of the viewer. Lacking a detailed, quantitative theory of human visual perception, we must content ourselves with a measure of image degradation, or the degradation that results when a codec is applied to an image. In the latter case, the problem is one of rating the similarity of two images.

The performance evaluation of different codecs is an important requirement. To this end, it is necessary to devise and implement suitable measures of performance. The performance of the cartoon-based cueing process can be evaluated using Receiver Operating Characteristics (ROC). In this approach, the percentage of true positive detection of real targets/regions of interests is assessed against the percentage of false negative detection of target areas as background. The targets/regions of interests are identified by the intelligence analyst and photographic interpreter and are used as ground truth in this project.

The first and most obvious metric for compression performance is "compression ratio", i.e. the reduction in the number of bits used to represent the image. The term compression ratio is useful in relation to lossless compression because it is concerned with the reduction in the number of bits used to represent all the data/information. In relation to lossy compression, however, compression ratio conveys no information regarding what information has been preserved and what data/information has been discarded. In general such a metric would need to be related to the relevant criteria for the task at hand.

In the case of the Content-Based Image Compression approach that we are developing, compression ratio is perhaps justifiably used as it is simple and it quantifies the removal of data defined (by content description) as redundant. Thus, the compression achieved depends on the overall dimension of all the regions/targets of interests in an image. Imagery with few small targets/regions of interests can be compressed significantly more than imagery with large areas/targets of interests. The achievable compression can be seen to be dependent on the amount of information that must be preserved losslessly in order to ensure value of the transmitted (compressed) image.

An algorithmic comparison can be made between the Content Based and conventional approaches using an analysis of the performance of the region cueing processes on images compressed by conventional lossy techniques. Using the annotated military imagery as ground truth, the results of the automatic cueing process with and without compression can be compared.

The "informative" value of the decompressed images will be assessed by the intelligence analyst and photographic interpreter involved in the project. Their assessment will be used for guiding the development of the Content Based compression approach. A Mean-Square-Error (MSE) or Peak Signal-to-Noise Ratio (PSNR) is used to measure the distortion introduced by the lossy system in the regions defined as informative and the distortion in the non-target areas. The latter is important as the surrounding parts of the image can also play a role by supplying contextual information. The assessment based on the MSE/PSNR and the photographic interpreter/intelligence analyst will be evaluated to determine if there is in fact a corresponding relationship between the objective and subjective measures.

Another important metric is computational load, or the time required to carry out the compression/decompression.

Performance Visualisation

Granted that there are many important dimensions of performance which must be taken into account when evaluating compression systems, the practical reduction of this complex space to single figures of merit quite simply destroys the necessary information. However, understanding and assimilating this complex space is a significant problem for the human and a multi-dimensional graphical representation is necessary. An interactive performance evaluation visualisation tool is therefore being developed to facilitate comparison of the performance of conventional and content-based compression approaches using, among others, the following criteria:

- Target cueing in raw/decompressed images;
- Information preservation;
- Compression ratio;
- Computational load;
- Mean-square-error/Peak Signal-to-noise ratio;
- Photographic interpreter and intelligence analysts assessment.

The work that we have carried out on visualisation is concerned with the presentation of the performance evaluation measures of the efficiency and effectiveness of different compression approaches at different compression ratios in different circumstances. It aims to facilitate the assessment and selection of appropriate compression approaches and compression ratios for the image/image type in question. It provides, for example, a means to understand the degree of invariance of the information cueing and preservation to the presence of a compression-decompression step. This is shown in Figure reffig:visual, which is a three-dimensional visualisation of codec performance. The three dimensions are: the codec used (horizontal, left), the type of imagery under investigation (horizontal, back), and the compression ratio achieved. The latter is shown as a vertical bar: green is positive (ie above the horizontal plane), whereas red shows negative (below the plane).

7 Evaluation of coder performance

7.1 Previous Work

Conventionally, image degradation is measured using the Mean Square Error (MSE) between the images. This is defined as the average of the pixel-wise square difference between two images $\mathbf{I}$ and $\tilde{\mathbf{I}}$:

$$MSE = \frac{1}{N} \sum_{x,y} (\tilde{I}_{x,y} - I_{x,y})^2 \tag{4}$$

The MSE has the virtue that it is easy to calculate. However, it is not a very good measure for many classes of degradation. It works well for unstructured degradations, such as the addition or suppression of uncorrelated noise. On the other hand, precisely because it is defined only on

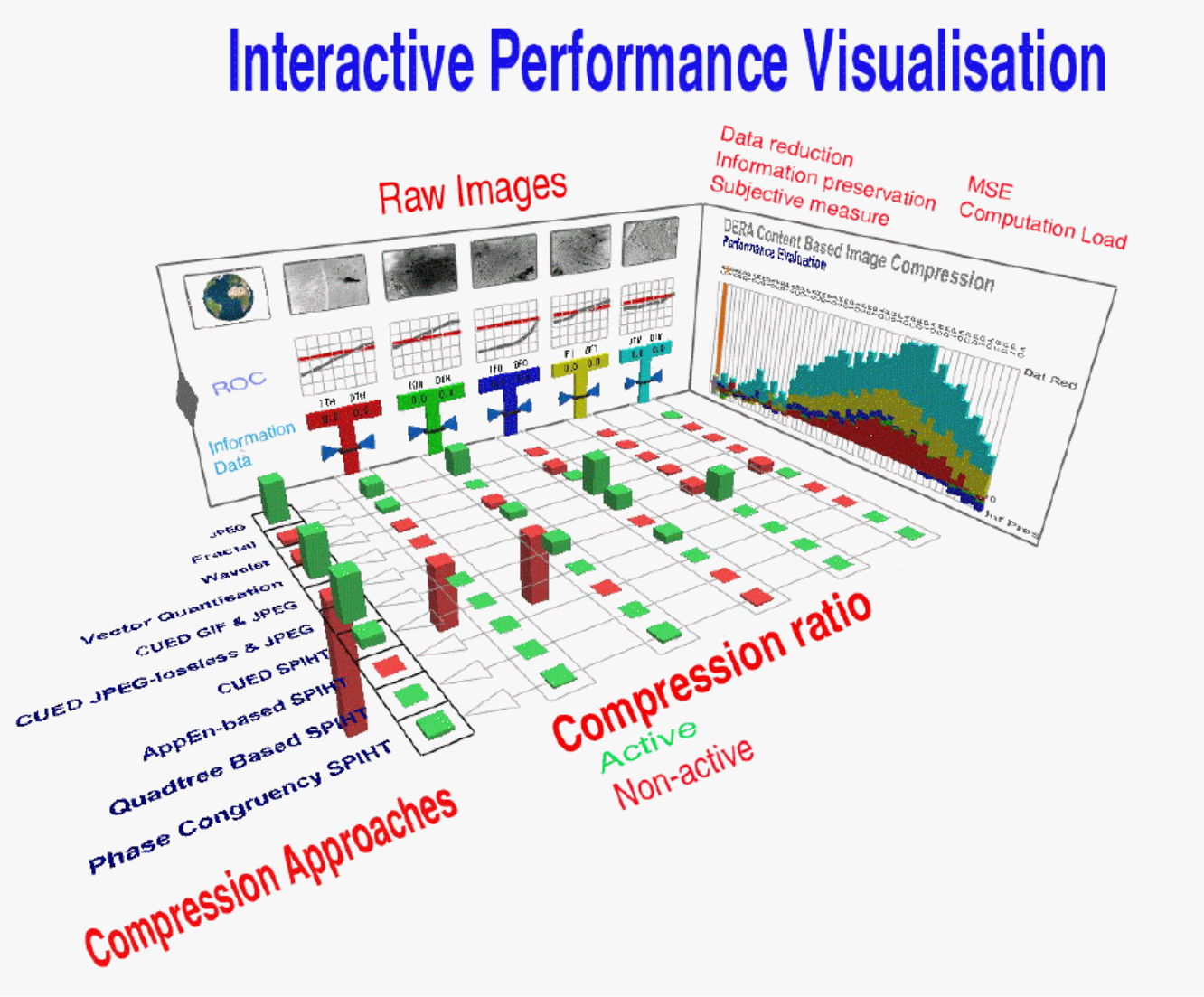


Figure 14: Visualisation Tool.

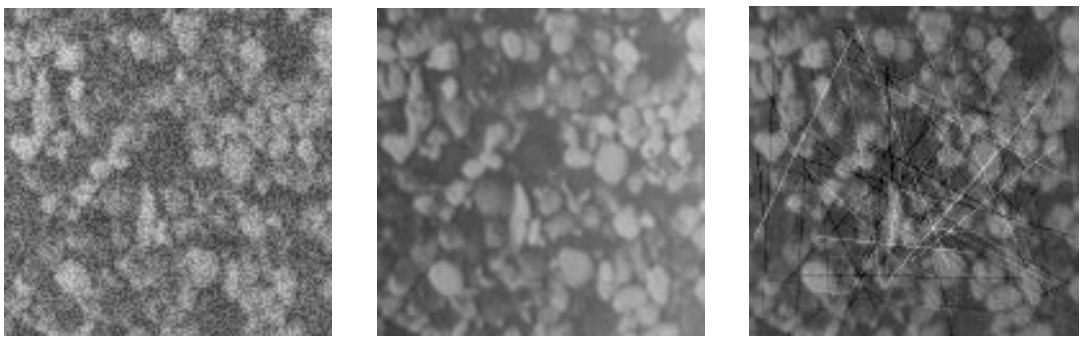


Figure 15: **Left** – *pebbles*+uncorrelated noise : MSE = 1204 **Centre** – *pebbles*+ intensity slope: MSE = 1200 **Right** – *pebbles*+ structured artifacts: MSE = 1200

the basis of individual pixel values, it cannot take account of correlated artifacts in a meaningful way. This is an unfortunate limitation since Section 5 highlighted the importance of preserving structured information. For example, Figure 15 shows three degraded version of the image *pebbles*, all of which have very similar MSEs. It can be seen that the spatial organisation of the degradation, which is not accounted for by MSE, determines the degree to which image features are obscured. The addition of the intensity slope (Figure 15: Centre) does not cause features to be obscured, while the addition of either uncorrelated noise or structured artifacts (Figure 15: Left, Right) can lead to misinterpretation. In the case of uncorrelated noise, this is solely due to masking of the original features, while the introduction of structured artifacts may lead to further confusion, as the artifacts may be mistaken for real image features. Figure 15 clearly demonstrates that MSE is not a suitable for measuring the damage caused by a general degradation process.

Despite this inherent limitation of MSE, it is still used overwhelmingly in practice. Several attempts have been made to overcome this limitation, for example [32], who weight the error at each image location according to a measure of the "visual importance" of the locality. However, all such measures are based on pixel-wise summation, and so they also cannot take account of correlation in the error.

Several attempts have also been made to create comparison schemes based on aspects of human perception, for example [13], [46], [56] and [31]. Typically, these incorporate findings about the earliest stages of human perception, mostly for the first few tens of milliseconds that an image is regarded. For reconnaissance or medical images, however, this limitation is not relevant, since the images may be studied at length, and the viewer may have access to such tools as zoom and contrast enhancement. In general, most of the measures that have been developed aim to give a good overall impression rather than on preserving the detail which may carry vital information.

7.2 General Paradigm

An image comparison method should account for changes in the informative content of the image. The task is one of:

1. Identify relevant image content;
2. Describe, or at least represent, this image content;

3. Assess preservation of the image content, by comparing the description of the degraded image to that of the original.

As we have noted at various junctures in this article, what constitutes image content depends ultimately on the particular application. In such cases, it is necessary to construct a semantic image model, which can express all relevant information in a meaningful form. The image can then be described by recording the model parameters.

Note that many images may have the same description. Indeed, we would like images with the same semantic meaning to have the same, or similar descriptions. The description need not contain all the information of the image, only that which is relevant to the application. It is important that information which is irrelevant to the application be ignored at the description stage, or by the function that measures the "distance" between a pair of images. For example, one of the (many) failings of MSE is that it responds strongly to a uniform intensity shift, though such a shift does not change the perceptual content of the image. However, if we know that the degradation processes to be observed do not affect certain image properties, then we need not be concerned about the dependence of the distance function on them. The widespread use of MSE can, to some extent, be justified by the fact that coding schemes are unlikely to result in a uniform intensity shift.

When comparing image descriptions, the relative significance of parameters must be considered. Significance may be pre-defined according to the image model and to the application, e.g. intensity may be more important than colour. Alternatively, it may be assigned according to the "rarity" of a parameter value within a given image. The significance of rarity is well established in the field of Information Theory, and this has recently been related to position coding in image analysis [15; 21]. Figure 16 is reproduced from [21] and demonstrates how, even at a casual glance, "rare" features contain disproportionately more information than those that occur frequently in the image.

To compare two images, we need to define a distance function, which can be applied to the vectors of image model parameters computed from the original and degraded images. This needs to reflect both the degree to which the image has been altered, and the saliency of the original image. It may also be necessary to include a measure of the saliency of the degraded image, since introducing a salient feature may also lead to misinterpretation.

To date, we have attempted to compare images at a low level, devoid of application-specific knowledge. As a result, we seek to create a measure that is applicable to a wide range of images and degradation processes.

7.3 Example One: A line based approach

A local description is useful for a region-based coding scheme. This allows integration of the degradation measure over arbitrary regions, and so enables flexible segmentation schemes. As we noted at the outset of this article, images may often be modelled as consisting of smooth regions, separated by, or interwoven with, singularities. It appears that (for grey level images) much of the important information is expressed by these singularities. Local Energy and Phase Congruency, as introduced in Section 2, is a local measure, which has associated with it properties which described such sigularities.

Phase Congruency also provides valuable information for describing the feature. The phase can indicate the type of feature: step, ridge, etc.. Also, a small translation of the feature results in a phase shift measured near the feature. Orientation and energy of the feature can also be

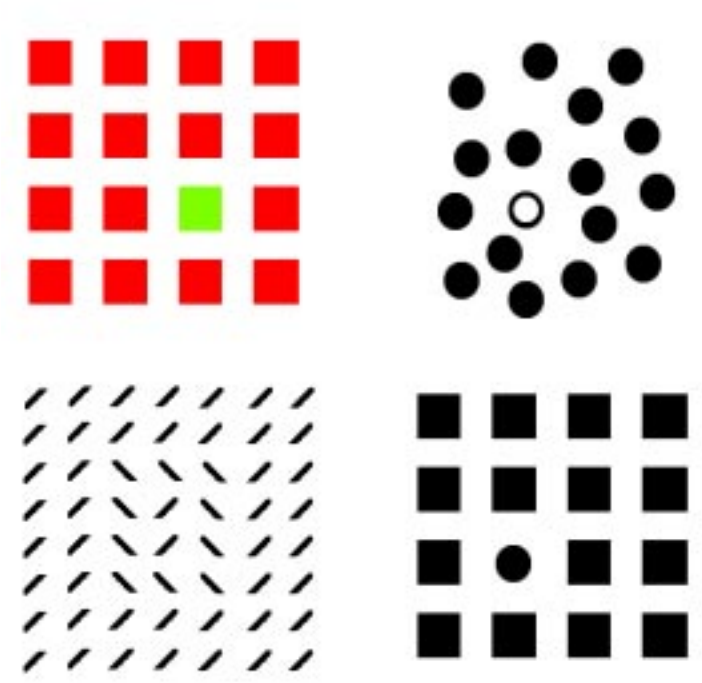


Figure 16: Demonstration of rarity and visual saliency: The eye is drawn to features which are, in some way, rare within the image.

computed. However, as noted in Section 2, care must be taken in the representation of sub-features. If properties such as phase are integrated over all scales, artifacts such as ringing are masked by stronger features near by. The resolution of these problems remains an open problem. Despite these problems, phase congruency can be used to measure image degradation, at least in cases where the underlying feature is grossly altered in terms of direction or phase.

We present two scenarios of how data generated by phase congruency could be used to construct an application-specific measure of information lost during processing/compression of an image.

Scenario 1: Validation of edge extraction

The aim is to devise a check for the validity of an edge map extracted from an image using the Canny edge detector [7]. In this particular case, we are not interested in the fact that we may have missed lines. Rather, we want to ensure, that all lines correspond to a linear feature in the original image. We first apply the Canny edge detector to *pebbles*. For each point which is determined to be part of a line, we compute the dominant direction, using a LogGabor decomposition over four scales, and compare these directions with those in the original image (computed in the same manner).

Figure 17 shows the original image *pebbles*, and figure 18 illustrates the error in the dominant direction. Black corresponds to points where the same direction was detected in both the edge map and the original image, whereas white corresponds to a large difference in direction. However, the figure shows that, on the whole, the edge detector produces lines which corespond to the dominant directions in the image. Deviations can be attributed to the fact that the

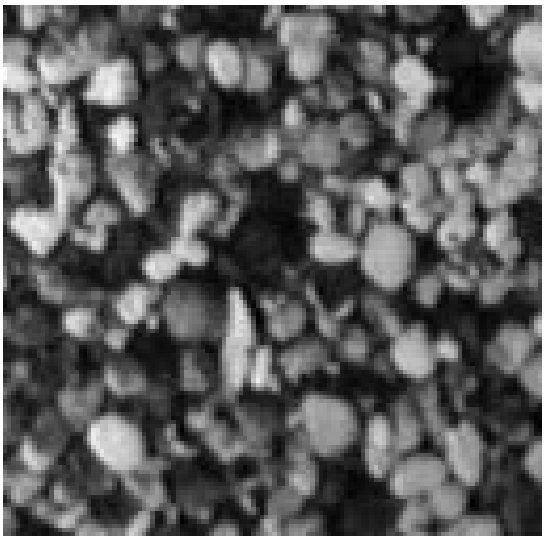


Figure 17: *pebbles*

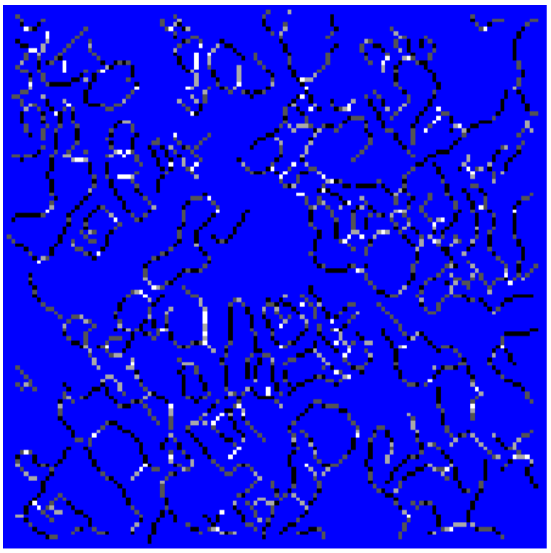


Figure 18: Deviation of line direction from dominant direction in original.

	JPEG Quality	Using Phase Conguency	Using Energy
pebbles	25	2.57	70.7
	50	2.02	52.3
wood	25	1.44	34.87
	50	1.19	26.7

line orientation is calulated at a range of scales, and as the pebbles are small and have curved boundaries, the orientation is ill-defined across scales. The largest deviations are generally associated with lines which do not correspond to clear isolated edges.

This is an example of a very general (and rather crude) description process, which can extract a compatible description from very different representations of information (The edge map and the original image). For this reason it does not take advantage of aditional information present in the original image, such as Local Energy.

Scenario 2: Strong degradation by JPEG

In the second scenario, we assess degradation caused by JPEG with a low quality setting. In this case, we require more than the correct direction. We compute any difference in the dominant direction, and then the differences in phase, at those locations where the dominant direction remains unchanged. The sum of these two error measures was weighted using either phase congruency or local energy as a saliency measure. This is a simple example of the image degradation paradigm: relevant information is identified using PC or Local Energy, and described in terms of local orientation and phase. The descriptions are compared using a simple metric, which was averaged across the whole image to give a numerical degradation value.

7.4 Example Two: Entropy Based Approach

The description process described in section 7.3 is based on phase congruency and its associated properties. This was motivated by the fact that phase congruency is a good local indicator of salient information. A degradation measure that has local support is preferable. In such a case, the automatic determination of those localities that appear to contain important information is key. Entropy is a natural candidate for measuring information, being central to information theory. However, Shannon entropy is only one formalisation of the intuitive concept of entropy, and it turns out that it is not an ideal measure for this application. Fundamentally, this is because it is based on the probability density function (pdf) of the image, and does not take spatial structure into account. This means that it is unable to distingish between noise, or uncorrelated texture, and more structured features. We noted in Section 5 that commonly used compression techniques fail in ways that are perceptually salient and structured. A substantial number of entropy measures exist, in addition to Shannon's. In section 7.5 three entropy measures are introduced and compared, and *Approximate Entropy* is chosen as a suitable measure of information.

The relationship between Information Theory and Image Analysis is only now beginning to be worked out. Gilles and Brady [15] show how an information theoretical perspective leads to position codes for salient information that is based on local entropy and which facilitates image matching and geometrical registration.

7.5 Comparison of Entropy Measures

We introduce three entropy measures, complementary to that introduced by Shannon. In section 7.5.1 the importance of these measures is explored, with relation to the image models that implicitly underlie the measures.

Spatial Entropy

Spatial entropy is similar to Shannon entropy, except that the probabilities used are not estimated from a histogram. Instead, the pixel values themselves are treated as probabilities. The spatial entropy of an image, consisting of a lattice of sites x_i and associated intensity values I_{x_i} is defined as:

$$\sum_i \frac{I_{x_i}}{\sum_i I_{x_i}} \log \left(\frac{I_{x_i}}{\sum_i I_{x_i}} \right) \tag{5}$$

Spectral Entropy

Spectral entropy, is similar to spatial entropy, except that the image first undergoes a linear transformation $\mathcal{F}$, and the values from the transformed image are used to approximate the probabilities. The transformation used is typically the Fourier transform, hence the name. Recently, the wavelet transform has also been used. Spectral entropy is defined as:

$$\sum_i J_{x_i} \log J_{x_i}, \text{ where } J = \mathcal{F}(I) \tag{6}$$

Approximate Entropy

Approximate entropy[42; 43; 44] (ApEn) differs from Shannon entropy in the following two ways.

- It involves probabilities of pixel groups, rather than of individual pixels.
- Although it is based on a pdf, this is not approximated with a histogram.

To define ApEn, let $\mathbf{u}$ be the sequence of real values to be evaluated, $\mathbf{u} = (u_1, \ldots, u_N)$. $\mathbf{x}_i(m) = (u_i, \ldots, u_{i+m-1})$, $1 \geq i \geq N_m, \in \mathsf{Z}$ is a string of length m taken from $\mathbf{u}$. N_m is the number of strings of length m which can be taken from the sequence $N_m = N - m + 1$.

Strings are compared using a distance metric, d.

$$d(\mathbf{x}_i(m), \mathbf{x}_j(m)) = \max_{p=0}^{m-1} |u_{i+p}, u_{j+p}| \tag{7}$$

The number of strings which match string i, is given by:

$$C_m^i = \frac{1}{N_m} \#\{j : 1 \leq j \leq N_m, d(\mathbf{x}_i(m), \mathbf{x}_j(m)) \leq r\}, \tag{8}$$

and the *Approximate String entropy* of strings of length m, ϕ_m, is defined as:

$$\phi_m = -\frac{1}{N_m} \sum_{i=0}^{N_m} \log C_m^i \tag{9}$$

Finally, ApEn is defined as:

$$ApEn = \phi_{m+1} - \phi_m \tag{10}$$

If values u_i are restricted to a finite set and if $r = 0$, then ϕ_m reduces to the *Shannon entropy* of strings of length m. ApEn measures the increase in entropy of strings as the $(m+1)^{th}$ value is appended to them. It is a measure of the uncertainty of a pixel's intensity, given that its neighbors are known. It is instructive to use a revised form of ApEn suggested by Rukhin [47], in which ApEn is computed on a cyclic version of the sequence to be tested. This allows $N_m = N_{m+1} = N$, where N is the length of the sequence.

$$ApEn = -\frac{1}{N} \sum_{i=0}^{N} \log \frac{C_{m+1}^i}{C_m^i} \tag{11}$$

If now $r = 0$ and the sequence consists of values from a finite set, ApEn reduces to the "Shannon Conditional entropy". This is the Shannon entropy based on the conditional probabilities of a value occurring in the sequence, given the preceding m values.

ApEn has one major advantage over Shannon entropy. It has similar sensitivity to the matching tolerance r, as Shannon entropy does to the bin size, but it is not affected by uniform shifts in intensity. On the other hand, Shannon entropy may fluctuate as the intensity shift is increased due to the arbitrary placing of bin boundaries. This effect is illustrated in Figure 19. This does not happen with Approximate entropy as the matching is based on distance between data points alone. All Histogram based entropy measures can be stabilized in this way, but only at the expense of computational efficiency.

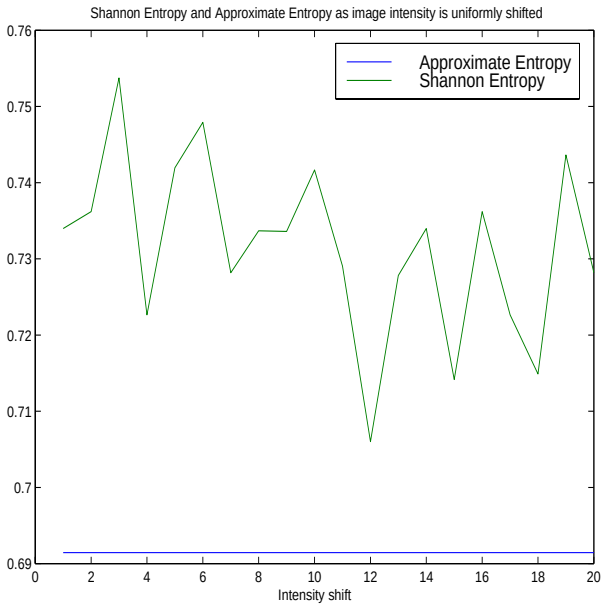


Figure 19: Comparison of the stability of Approximate Entropy and Shannon entropy to uniform shifts in image intensity

7.5.1 Implicit Models of the Entropy Measures

Entropy measures are based on some underlying "prior model", which corresponds to the minimum-entropy case. A high entropy value results from a data set which does not fit the prior model well. To identify this model, we must consider entropy in terms of information.

There is some confusion in the literature over the relationship between information and entropy, particularly Shannon entropy. Usually, if a data set has low Shannon entropy, it is associated with high information. Clearly, this is not satisfactory for the current application, as the minimum-entropy image is monotone, and is of little interest. What is meant is that the *pdf*, on which the entropy is based, incorporates a great deal of information about the image. Indeed, when the Shannon entropy is zero, the image is uniquely specified by the pdf. Moreover, it is precisely the fact that it can be so easily represented which makes it uninteresting. Images consist largely of uniform regions divided by boundaries. It is these boundaries which are of most interest, and which are of most significance to the human visual system. The minimum-entropy image for Shannon entropy corresponds well to uninteresting image regions. The more remote the image is from this case, the less information the pdf provides, as there are more possible pixel configurations compatible with it. For the purposes of this paper, we may say that the image contains more information, meaning more information not attainable from the pdf.

For Spatial Entropy the minimum-entropy image is that for which only one pixel is nonzero. This image provides maximum information about the behavior of photons which produced the image. This model is not as intuitively useful as that for Shannon entropy. Nevertheless, spatial entropy has other merits that we discuss later. The maximum-entropy image for spatial entropy is the monotone image, and so low values of spatial entropy indicate saliency.

With ApEn, the prior model is complete predictability of one pixel given another; this is only possible for the monotone image (or one that is striped/checked). The set of images which give low ApEn are those with a limited number of (nearly) monotone regions. This is a

stricter condition than required to give low Shannon entropy which is insensitive to the spatial configuration of the pixels.

The minimum-entropy image for spectral entropy, a sinusoid, is not so obviously useful for image analysis. When dealing with this measure, the minimum-entropy image is only a good background model if very low frequency cosines are considered. In real images, low spectral entropy usually indicates that only low frequencies are present, and this does fit our notion of background well. Even exceptional regions cannot be expected to present all frequencies with equal energy. However, as high frequency elements are introduced, the spectrum does become flatter, and entropy is increased. The measure is an indication of the spectral decay at high frequencies, and is related to the Lipschitz regularity.

7.5.2 Entropy and Saliency

The prior model for Shannon entropy, and for ApEn, is the monotone image. For Spectral entropy, it could be said to be a smooth image, while for the spatial entropy it is a delta-function. It follows that entropy measures how appropriate this model is for the data. Each model has one or more degrees of freedom. For example, with Shannon entropy the intensity of the monotone image does not matter. The entropy is a measure of the statistical uncertainty in this parameter. A high uncertainty indicates that the model is not well suited to the data.

In the case of Shannon entropy, high entropy implies high saliency, as we are interested in regions where the model does not fit well. However, when spatial entropy and ApEn are used, it is low values which correspond to salient regions. In these two cases, the minimum-entropy model is of little relevance; but from the images likely to be encountered, those with low entropy are likely to be more salient that those with high entropy. As long as ApEn is calculated with a small matching tolerance r, it is only low in areas of structure and not in background regions, as in practice there is likely to be sufficient noise/texture in these regions that they are not predictable.

The underlying model behind Shannon entropy makes it an obvious choice for a saliency detector. However, for images that are quite noisy or textured, it may respond as strongly to texture as to structure, and ApEn often gives better results. As images are, in practice, quite often noisy or textured, ApEn is a clear choice for detecting structure in the image. However, it is only a good choice, so long as we are not likely to encounter a monotone image or region.

7.6 Example of ApEn as a saliency cue

Figure 20 shows an example of using ApEn to cue regions, discarding a percentage of the least significant coefficients in the unwanted areas. The resulting image was then coded with the SPIHT coder, and finally decompressed at different bandwidths. The particular bandwidths used were: 290, 130, 60, 40, 20, and 10:1. Note that at the most severe compression (290:1), detail is lost in key areas; but that this is restored far more quickly (eg at 60:1) than the background.

7.7 Example of Entropy as a measure of image quality

In this section an ApEn based measure is used to detect the presence of artifacts in an error image, created by subtracting the degraded image from the original. This does not fulfill the criterion of a good degradation measure set out in Section 7.2, as it does not include any notion

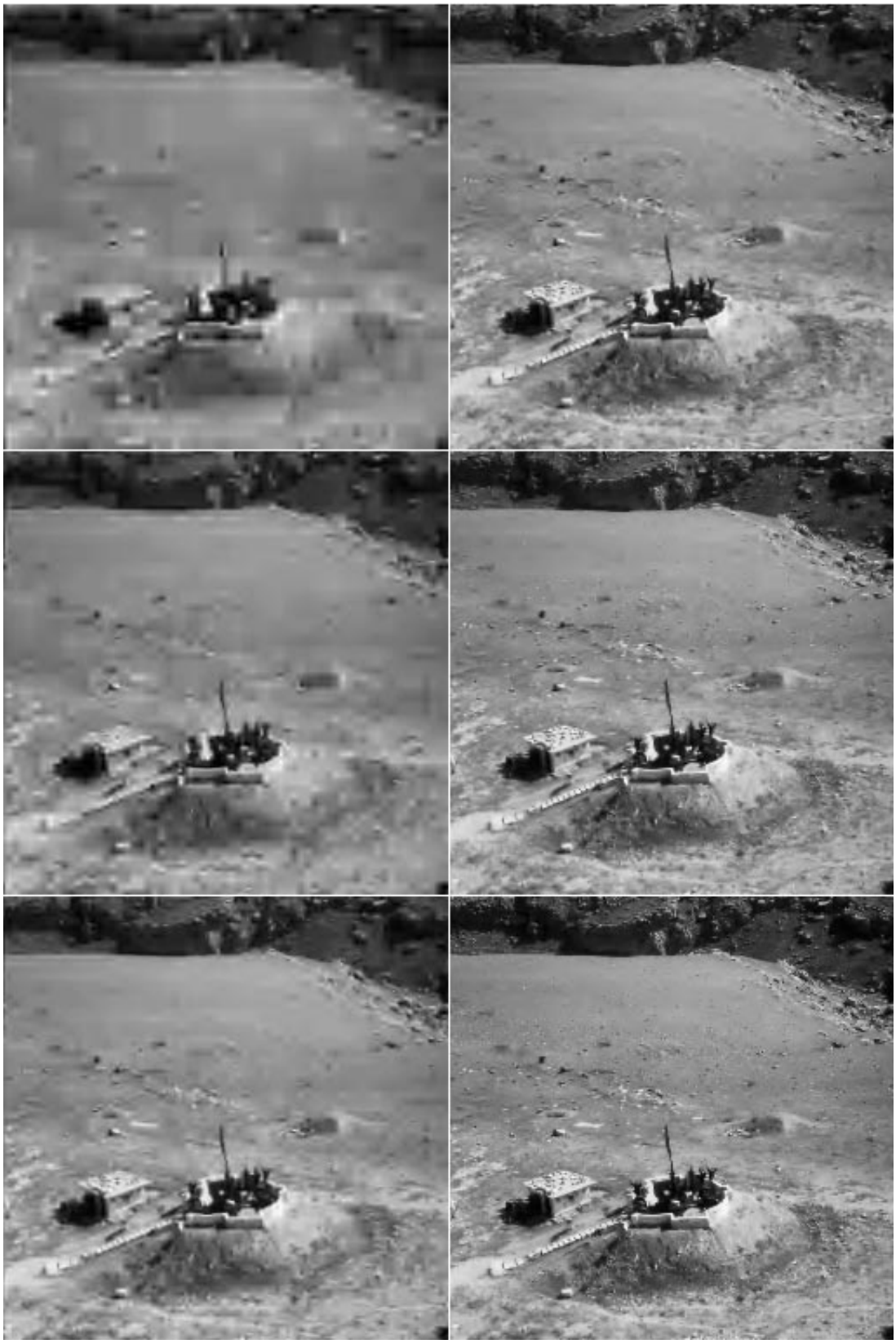


Figure 20: Coding scheme driven by ApEn

of the information present in the original image. Nevertheless, the measure that we develop performs significantly better than MSE.

The method involves applying ApEn locally to the error image, which is constructed by taking the pixel-wise difference between the original and the reconstructed images. This produces an "ApEn map" which may be averaged to give a measure of the presence of features in the error image.

Low ApEn corresponds to structured features in the error image, which correspond to a loss of image features, or to the introduction of artifacts. However, ApEn is not suitable for use with high quality image reproduction, as a monotone error image, corresponding to perfect reconstruction, also results in low ApEn. In order to compensate for this, ApEn can be weighted with the Shannon entropy, which provides a good measure of "how much is going on" locally [17].

If the matching tolerance $r = 0$ and the string length $m = 1$ then $\phi(m)$ is the Shannon entropy and the weighting can be worked into the ApEn calculation to create a new "structure function" S.

$$S = \sqrt{\phi_{m+1}^2 + k} - \phi_m \tag{12}$$

Note that if a circular or augmented window is used, $0 \leq \phi_m \leq \phi_{m+1}$. Lines of constant S are hyperbolae, so that, away from the origin, $S \approx ApEn$, but where ϕ_m is small, $S > ApEn$. Low values of S indicate that there is a wide spread of intensities present, but that the value of a pixel may be predicted with high certainty given the value of the preceding pixel, and hence to significant structure in the error image. The images 21-24 have been compressed using JPEG, with four different quality settings. Information is measured in the error image, using pixel-wise Mean Square Error, MSE, Shannon entropy, ApEn and the Structure function described above. The entropy measures were computed using a sliding 7x7 window, and the results were averaged to give a numerical evaluation of image quality, Table 7.7.

The degradation measure is $-\log S$. The monotone image needs only the DC component to describe it, and so no degradation is present. In the random image the degradation is considered small for all qualities, as the noise contains no structure. The 25 and 50 measures are a little higher, because of blocking. It can be seen that the degradation measure based on ApEn is able to distinguish between the removal of noise and of perceptual features, whereas Shannon Entropy and MSE cannot. ApEn is able to detect structure wherever it occurs and give an objective measure of "how much" structure is present. No a-priori information about the coding scheme is necessary.

Even if the statistical process which introduced the error is an independently identically distributed (iid) random variable, images locations where this random variable happens to cause structural degradation are detected. This would be unhelpful if the noise introduced was different each time, but coders usually give repeatable results on a given image[2] and so this is appropriate.

Note that the measure does not depend strongly on image contrast. The structure function can detect more varied structure than methods based on autocorrelation. It does not require that neighboring pixels have similar intensities, only that pixel intensities are predictable given their neighboring intensities. The main disadvantage of the structure function is that it is very sensitive to the parameter k.

[2] Only the encoding and reconstruction process are under consideration here, not noise introduced in transmission

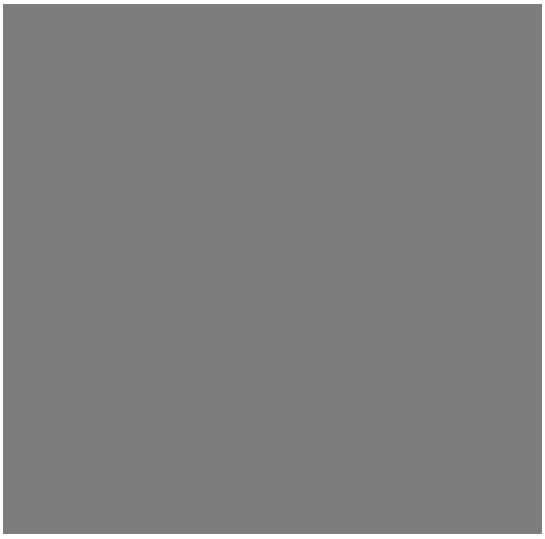


Figure 21: Monotone

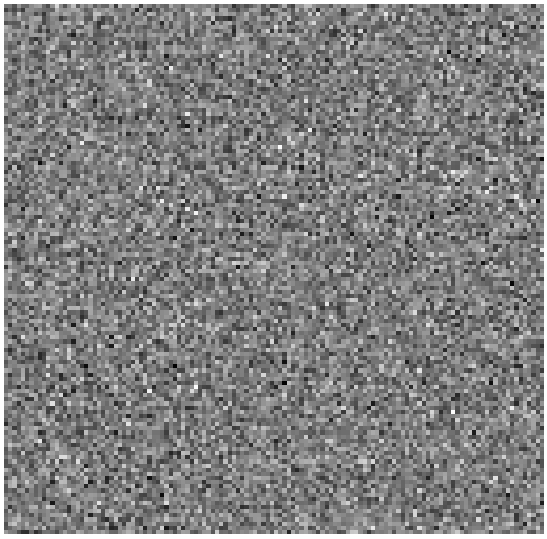


Figure 22: Random: uniformly distributed noise, with a range of 70 gray-levels.

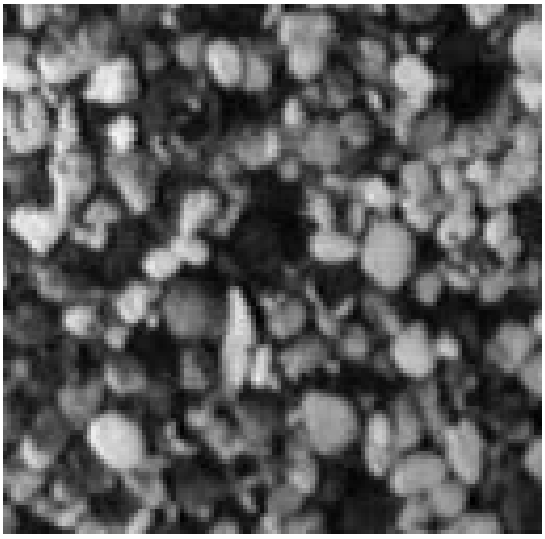


Figure 23: Pebbles

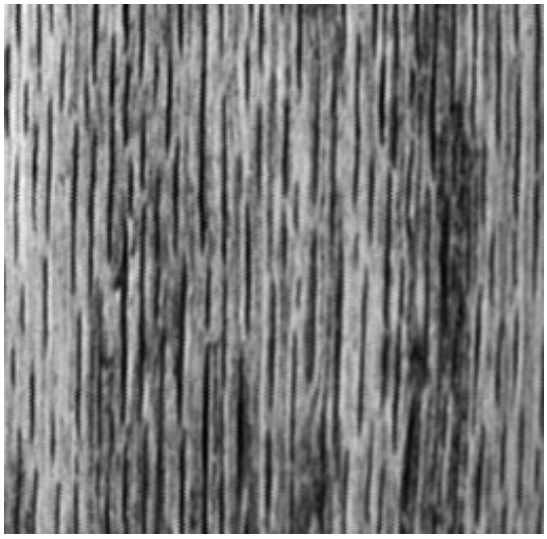


Figure 24: Wood-Grain

	Quality	Shannon Entropy	ApEn $\times 10^{-}1$	Structure $\times 10^{-}1$	Degradation	MSE $\times 10^{+}1$
MONOTONE	10	0	0	1	0	0
	25	0	0	1	0	0
	50	0	0	1	0	0
	75	0	0	1	0	0
RANDOM	10	0.6608	0.4758	0.8872	0.05199	4.121
	25	0.7815	0.4856	0.8758	0.0576	263.8
	50	1.065	0.5331	0.8436	0.07388	280
	75	0.6428	0.4758	0.8872	0.05199	0.8022
PEBBLES	10	1.382	0.3335	0.6126	0.2128	88.96
	25	1.293	0.429	0.7098	0.1489	34.09
	50	1.2	0.4969	0.781	0.1073	17.86
	75	1.141	0.5485	0.8385	0.07647	10.45
WOOD-GRAIN	10	1.408	0.303	0.5802	0.2364	114.6
	25	1.269	0.3872	0.6652	0.177	47.13
	50	1.181	0.4583	0.7393	0.1312	25.6
	75	1.209	0.5124	0.7979	0.09806	15.54

7.8 Multiscale Entropy

Whether we are using entropy as a saliency cue or to measure degradation, both the bin size and the window size must be set according to the features of interest. If the window is too small, representative statistics can not be obtained. On the other hand, if the window is too large then the statistics will not be sufficiently localized to represent the smaller features. Too much background will be included in the window, which can overwhelm the information from the feature.

Of course, the interesting structures in images typically occur at more than a single spatial scale in an image. This is particularly the case when there is substantial depth variation within the image. This necessitates a multi-scale approach, which can be achieved by measuring entropy over of series of windows of different sizes. There are, however, a number of problems with this method. In particular, similar objects are likely to be assigned a different saliency measure, depending on whether they occur in the foreground or the background. This is because the large window enclosing the foreground object includes detail which is not discernible in the background version due to the restricted resolution. We would prefer to identify similar saliency irrespective of whether the object is in the background or the foreground, and identify smaller "sub-features" at their own scales. This yields a measure of entropy which corresponds better to real life objects.

To estimate the "scale(s)" of a feature, we imitate the process of retreating from the object, by smoothing away detail using a low-pass filter. We can build a scale-space representation of the image, consisting of a set of images derived from the original by blurring. Each image is blurred to a greater degree than the previous. This gives a band-pass effect: the low-pass filter sets a lower bound on the size of the features of interest, while the window size sets an upper bound. Entropy can then be calculated from the scale stack to identify the locations and scales of the salient features. This sort of "scale-space" was introduced by Witkin **??**, who used a Gaussian filter. As the blurred images are smooth, they may be decimated creating a *scale pyramid*.

We now discuss some entropy measures based on Gaussian scale-space and Wavelets, and note the relationship between some of these methods to Approximate entropy.

7.8.1 Literature

Spectral entropy is one example of how multi-scale concepts can be incorporated into a complexity measure. As well as the Fourier domain, other bases may be used which are localized in space as well as in the Fourier domain. Of particular note are measures based on wavelets. These can take account of pixel correlation, unlike Shannon entropy which assumes pixels are uncorrelated and is well described by a pdf.

Murtagh's entropy measure [50] is based on wavelet coefficients, but the probabilities used are not the coefficients themselves. Instead the measure is based on a noise model and information is assigned in part to a signal term and in part to noise. This provides a useful tool for feature detection or image restoration in noisy conditions, where the noise model is known or can be estimated.

Wavelet based measures may be calculated across all scales over a given locality, or over all coefficients from one scale. However, the latter does not always give a clear description of which scales are most salient. This is because the wavelets form a sparse representation, and unless the location, scale and shape of the wavelet correspond well to the feature, the feature is represented by a large number of wavelets at different scales. Also, artifacts may be created according to the shape of the wavelet. This can only be avoided with a Gaussian kernel [57], which does not satisfy the criteria for a wavelet.

The work of Winter et. al. and of Jägersand adopts a different approach. A Gaussian scale-space is used and, in contrast to wavelet measures, which amount to calculating entropy on a number of band-pass images from a scale-pyramid, an attempt is made to compare the information present at one scale, with that at the next. A low-pass scale stack is used and the idea is to identify those scales and locations at which detail appears as the frequency threshold is increased. The scale-space is full resolution and so the measures are relatively stable with respect to translation of features.

Both methods use *relative* entropies. In the case of Winter, mutual information (MI) is used. If X and Y are two consecutive images from the scale stack, and their pdf are $P_X(i)$ and $P_Y(i)$, then:

$$MI = -\log \frac{P_X(i)P_Y(j)}{P_{X,Y}(i,j)} \tag{13}$$

Jägersand's method is based on the Kullback Contrast, $K[X,Y]$ between the spatial distribution of the consecutive images X and Y from the scale-stack.

$$K[X,Y] = \sum_i X_i \log \frac{X_i}{Y_i} \tag{14}$$

This was initially calculated globally, then on a patch-wise basis. It was compared to use of the Fourier domain. Identifying relevant scales in the Fourier domain was very difficult, while the they could be easily identified with the use of Kullback contrast. Features are not well localised in the Fourier Domain as sinusoids do not represent their shape well.

7.8.2 Multiscale ApEn-like measures

It has already been noted that ApEn is particularly good measure of saliency: it makes use of an appropriate background model, and is able to take account of structure. A problem with ApEn is that it is defined at a pixel-wise scale, so that high frequency noise may hide larger scale structures. We now explain how ApEn may be extended to detect structure at a range of scales.

Unless otherwise stated, ApEn is assumed to be calculated with m=1 and r=0. This is primarily in order to simplify notation and to allow equalities where only similarities exist for the general case. ApEn of a discrete signal A is now equal to the Shannon Conditional entropy between A and B, where B is a shifted version of A.

$$ApEn = -\sum_{i=0}^{N} \log P(B_i|A_i) \tag{15}$$

Any pixel pair represented by (A_i, B_i), can be represented equally well by the rolling average, C_i, and the rolling difference, D_i. C and D are generated as the high-pass and low-pass images, at the first step of a Haar wavelet transform and so an ApEn-like measure could be calculated on C and D at each step of the wavelet transform giving a multiscale measure. We consider taking the conditional entropy of C given D:

$$\begin{aligned}
H_c(D, C) &= -\sum_{i=0}^{N} \log P(D_i|C_i) && (16)\\
&= -\sum_{i=0}^{N} (\log P(D_i, C_i) - \log P(C_i)) && (17)\\
&= -\sum_{i=0}^{N} (\log P(A_i, B_i) - \log P(C_i)) && (18)\\
&= -\sum_{i=0}^{N} \log P(A_i) + \log P(B_i|A_i) - \log P(C_i) && (19)\\
&= ApEn - \sum_{i=0}^{N} \log \frac{P(A_i)}{P(C_i)} && (20)
\end{aligned}$$

The new conditional entropy differs from ApEn by a correction term, because there is not a one-one mapping from A to C. This term can be quite significant in the presence of features. A significance map generated from $H(D, C)$ is shown in Figure 26, and ApEn is illustrated in Figure 25, for comparison.

The second term may be eliminated by looking at $II = H_c(D, C) + H_c(C, D)$ which is identical to $H_c(A, B) + H_c(B, A)$, i.e. ApEn to the right plus ApEn to the left. This can be rearrange as

$$\begin{aligned}
II = H_c(A, B) + H_c(B, A) &= 2H(A, B) - H(A) - H(B)\\
&= H(A, B) - MI
\end{aligned} \tag{21}$$

Figure 25: Approximate Entropy, m=1, r=0

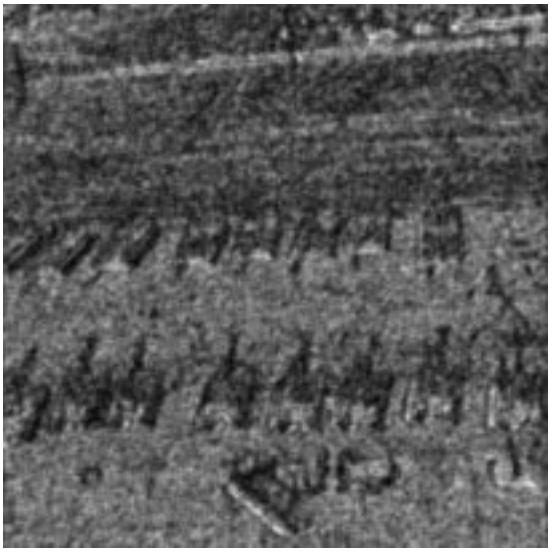


Figure 26: Conditional Entropy of High-Pass given Low-Pass

MI is the *Mutual Information*, a measure of how much information is shared between A and B. II is then the total information minus the mutual information, and represents the *Independent Information* of A and B. It can be calculated equally from A and B or from C and D, and so a multi-scale version can be calculated using the approximation and detail at each scale of a wavelet transform.

The use of II to detect information between scales is similar to the methods used by Winter and Jägersand, and provides a link between these methods and the use of ApEn on images. Both Winter and Jägersand compared consecutive levels in a low-pass scale stack, while we have compared the low-pass approximation with band-pass detail. Results are similar as the band-pass is simply the difference between the two low-pass images, and this choice eases the use of a scale pyramid as the high-pass and low-pass data is of the same resolution.

We have been lead to consider II, as an alternative to ApEn which better fits into the wavelet framework. Interestingly, Swelden's "second generation wavelets" [51] provide a multi-scale framework in which ApEn can be calculated directly at different scales. Second generation wavelets differ are not based on a translated and dilated mother wavelet as in the case of first generation wavelets. A typical method would be to take all even pixels as the approximation and predict the odd pixels using linear interpolation. The detail would then be the error in the prediction of the odd pixels. The simplest second generation wavelets are referred to by Sweldens as "lazy wavelets". In the lazy wavelet method, the image is simply decimated to give the approximation, and the discarded (odd) pixels constitute the detail. ApEn can clearly be calculated in the lazy wavelet framework, and would measure pixel predictability at varying distances. However, the practical use of such a measure on images has not yet been investigated.

7.9 Experiments with multi-scale entropy measures

The following experiments demonstrate the potential of using II to drive a compression scheme. However, some pitfalls are also identified and problems with saliency measures based on the wavelet transform are illustrated.

Figure 27: The original Image

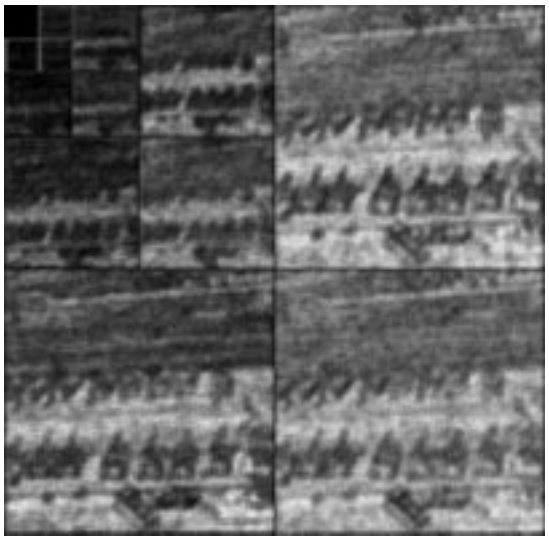


Figure 28: II map: bottom-left – II between approximation and horizontal detail at the first stage of the WT.

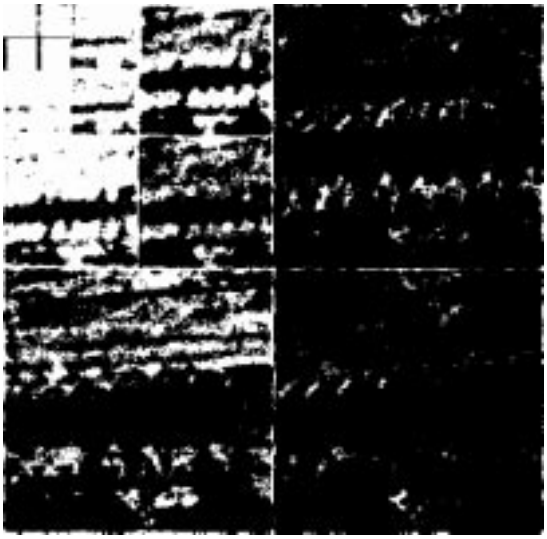


Figure 29: Threshold on 28 showing which coefficients are to be set to zero.

Figure 30: Image reconstructed from the 20% lowest II wavelets

Following the ApEn approach, we look for regions of low II. In many images this gives good results. E.g. Figures 27 - 30 demonstrate the identification of low II regions. Figure 27 is the original *tanks* image. Figure 28 shows II map, and Figure 29 shows which wavelet coefficients are to be kept. The rest are set to zero before computing the inverse wavelet transform in order to obtain the reconstructed image Figure 30. Although a better reconstruction can be achieved simply using the 20% largest coefficients, the reconstruction demonstrates that most of the important wavelets have been kept.

Although this method performs well on *tanks*, it performs less well on images with fine

structures such as Figure 1. The underlying asumption is that it a region of interest, activity at one scale, should occur at the some locations as activity at the scale below (coarser). This is not necessarily the case, as features only exist over a finite range of scales. Fine features, in particular, are not represented at the coarser scales in the decomposition.

8 Conclusions

Our goal has been to outline the design of an algorithm that achieves impressive levels of image compression without suppressing information that is key to decision making based on the transmitted image. At heart, the approach is based on integrating ideas from image analysis (computer vision) and image coding, two fields which have, to a surprising extent, developed relatively independently of each other. Our approach to date has been based on detecting significant features in images, specifically (open) curves and (closed) regions. In the case of intensity feature detection, we presented a number of developments that we have made to the local energy and phase congruency. Similarly, in the case of region segmentation, we argued for a texture model based on wavelet local energy and a development of the region competition algorithm that uses non-parametric statistics to evaluate the statistical force. The outputs of the curve and region algorithms are fused to develop a representation of the image that is submitted for coding. We have proposed a formal relationship between information theory and position codes for an image, which leads to an analysis of position codes that we call Unique Local Signatures based on local entropy. We use this to evaluate which coding algorithm is best suited to coding each region. Our initial experiments give very promising results.

Of course, this is merely a start to what is inevitably an ambitious project, and there are many open problems. First, there remains much work to do on feature detection, not least on automating the selection of scales at which features of interest exist, and separating out the confounding effects of nearby features. The relationship between our wavelet local energy and phase feature of textures is related to that for curves; but there are aspects of the relationship that require elucidation. The algorithm to choose an optimal coding scheme for each region is preliminary, as is its link to our theory of position coding. The current scheme is relatively slow and would need considerable work to make into a real-time practical system. To date, we have concentrated on single images, whereas in many cases data is from a moving camera. We have also concentrated on "low level" representations of an image and have only recently begun to study the ways in which more application-specific "higher level" knowledge can be integrated with the current approach.

9 Acknowledgements

This work was carried out as part of the MOD Cooperate Research Programme *Technology Group 10* as a collaboration between Oxford University and DERA.

References

[1] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. on PAMI*, 16(6), 1994.

[2] Dana H. Ballard and Christopher M. Brown. *Computer Vision*. Prentice Hall, 1982.

[3] A. Blake and M. Isard. *Active Contours*. Springer Verlag, 1998.

[4] R. Bracewell. *The Fourier Transform and its Applications.* McGraw Hill Book Company, New York, 1986.

[5] Michael Brady, Fuxing Li, and Zhiyan Xie. Texture segmentation from region competition and wavelet local energy. 2000.

[6] D. C. Burr and M. C. Morrone. Feature detection in human vision. *Proc. R. Soc. Lond. B*, 235:221–245, 1988.

[7] J. F. Canny. A computational approach to edge detection. 8(6):112–131, 1986.

[8] R. J. Clarke. *Digital compression of still images and video.* Academic Press, New York, 1995.

[9] I. Daubechies. *Wavelets.* SIAM, Philadelphia, 1992.

[10] Rachid Deriche. Using canny's criteria to derive and optimal edge detector recursively implemented. 1:167–187, 1987.

[11] P. G. Ducksbury. Parallel texture segmentation using a pearl bayes network. In *British Machine Vision Conference, BMVC-93, University of Surrey*, pages 187 – 196, 1993.

[12] P. G. Ducksbury, D. M. Booth, and C. J. Radford. Improved vehicle detection in infrared linescan imagery using belief networks. pages 415–419, 1995.

[13] P. Franti, T. Kaukoranta, and O. Nevalainen. Blockwise distortion measure in image compression. *SPIE*, 2663:78–87, 1996.

[14] D. Geman, S. Geman, and P. Dong. Boundary detection by constrainted optimization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(7):609–628, 1990.

[15] S. Gilles and Michael Brady. Information theory in computer vision: a sound framework for matching and recognition. 2000.

[16] Sebastien Gilles. *Robust description and matching of images.* PhD thesis, Oxford University, 1998.

[17] Sebastien Gilles. *Robust Description and Matching of Images.* PhD thesis, Department of Engineering Science, University of Oxford, 1998.

[18] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conference*, pages 188–192, 1988.

[19] G. Held. *Data and Image Compression : Tools and Techniques (4th Edition).* John Wiley, New York, 1996.

[20] R. P. Highnam and Michael Brady. *Mammographic Image Processing.* Kluwer series in computer vision, 1999.

[21] T. Kadir and Michael Brady. Saliency, scale, and image description. *Int. J. Computer Vision (submitted for publication)*, 2000.

[22] T. Kanungo, B. Dom, W. Niblack, and D. Steele. A fast algorithm for mdl-based multi-band image segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, 1994.

[23] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. In *Proc. 1st International Conference on Computer Vision*, pages 259–268, London, 1987.

[24] P. Kovesi. Invariant measures of image features from phase information. *Videre*, 1, 1996.

[25] Y. G. Leclerc. Constructing simple stable descriptions for image partitioning. *The International Journal of Computer Vision*, 3(1):73–102, 1989.

[26] Bernard Liang and Michael Brady. Feature descriptions based on phase congruency. *Image and Vision Computing (submitted)*, 2000.

[27] Tony Lindeberg. Edge detection and ridge detection with automatic scale selection. In *CVPR'96 : IEEE Conf on Computer Vision and Patten Recognition*, pages 465–470, San Francisco, California, June 1996. IEEE Computer Society Press.

[28] Tony Lindeberg. Automatic scale selection as a pre-processing stage for interpreting the visual world. *Proc. Fundamental Structural Properties in Image and Pattern Analysis FSPIPA 99*, 130:9–23, 1999.

[29] Stephane Mallat. *A wavelet tour of signal processing*. Academic Press, New York, 1998.

[30] D. Marr. *Vision*. W. H. Freeman, San Francisco, 1982.

[31] M. Miyahara, K. Kotani, and V. R. Algazi. Objective picure quality scale (pqs) for image coding. *IEEE trans. on commun.*, 46(9), 1998.

[32] John Morton and Ed Jernigan. Perceptually based image comparison. In *ICIP*, volume 1, pages 489–493, 2000.

[33] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, 42:577–684, 1989.

[34] M. Nielsen, L. Florack, and R. Deriche. Regularisation, scale space and edge detection filters. 7 (4):291–308, 1997.

[35] Alan V. Oppenheim and Jae S. Lim. The importance of phase in signals. 69:529–541, 1981.

[36] R. A. Owens and M. C. Morrone. Feature detection from local energy. *Pattern Recognition Letters*, 6:303–313, 1987.

[37] Theo Pavlidis. *Algorithms for graphics and image processing*. Computer Science Press, 1982.

[38] D. E. Pearson. Visual communication systems for the deaf. COM-29:1986–1992, 1981.

[39] D. E. Pearson. Transmitting deaf sign language over the telecommunications network. 20:299–305, 1986.

[40] D. E. Pearson and J. A. Robinson. Visual communication at very low data rates. 73 (4):795–812, 1985.

[41] P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. 12(7):629–639, 1990.

[42] S. M. Pincus. Approximate entropy as a measure of system compexity. *Proc. Natl. Acad. Sci. USA*, 88:2297–2301, 1991.

[43] S. M. Pincus. Approximate entropy: statistical properties and applications. *Commun. Statist.-Theory Meth.*, 21:3061–3077, 1992.

[44] S. M. Pincus. Not all (possibly) "random" sequences are created equal. *Proc. Natl. Acad. Sci. USA*, 94:3513–3518, 1997.

[45] M. Rabbani and P. W. Jones. *Digital Image Compression Techniques*, volume TT 7. SPIE Tutorial texts in Optical Engineering, 1991.

[46] A. M. Rohaly, A. J. Ahumada, and A. B. Watson. A comparision of image quality models and metrics predicting object detection. *SID Digest of Technical Papers*, (6.2), 1995.

[47] A. L. Rukhin. Approximate entropy for testing randomness. Technical report, Department of Mathematics and Statistics, University of Maryland, 1998.

[48] J. M. Shapiro. Embedded image coding using zerotrees of wavelets coefficients. 41:3445–3462, 1993.

[49] S.M. Smith and Michael Brady. Susan - a new approach to low level image processing. *International Journal of Computer Vision*, 23(1):45–78, 1997.

[50] J.L. Starck, F. Murtagh, and F. Bonnarel. Multiscale entropy for semantic description of images and signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.

[51] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.

[52] John Taylor. Engineering the information age. *Engineering Management Journal*, 1998.

[53] M. J. Varga and P. G. Ducksbury. Infrared thermography. 1997.

[54] S. Venkatesh and R.A. Owens. On the classification of image features. *Pattern Recognition Letters*, 11:339–349, 1990.

[55] Han Wang and Michael Brady. Real-time corner detection algorithm for motion estimation. *Image and Vision Computing*, 13(9):695–703, 1995.

[56] A. B. Watson, R. Borthwick, and M. Taylor. Image quality and entropy masking. *SPIE Proceedings*, 3016(1), 1997.

[57] A. P. Witkin. Scale-space filtering. In M. A. Fischler and O. Firschein, editors, *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, pages 329–332. Kaufmann, Los Altos, CA., 1987.

[58] Zhi-Yan Xie and Michael Brady. Texture segmentation using local energy in wavelet scale space. 1996.

[59] Zhiyan Xie and Michael Brady. Lacunarity: a fractal-based texture feature for natural texture segmentation. 1993.

[60] G. Xu, E. Segawa, and S. Tsuji. Robust active contours with insensitive parameters. *Pattern Recognition*, 27(7):879–884, 1994.

[61] S. C. Zhu and A. L. Yuille. Region competition: unifying snakes, region growing, and bayes/mdl for multiband image segmentation. *IEEE Trans. Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996.