

FEATURE DETECTION AND FUSION FOR INTELLIGENT COMPRESSION

P. G. Ducksbury[‡], M.J.Varga[‡]

ABSTRACT

In previous work [1] a novel approach was described which used automatic target detection together with compression techniques to achieve intelligent compression by exploiting knowledge of the image content. In this paper a set of standard feature detectors such as HV-quadtrees, approximate entropy and phase congruency are used as target discriminators. These detectors all attempt to find potential areas of interest within an image but will undoubtedly be slightly different in their estimates. A probabilistic (*Bayesian belief*) network is then used to fuse this information into a single hypothesis of 'interesting areas' within an image. A wavelet-based decomposition can then be applied to the image in which selective destruction of wavelet coefficients is performed outside the cued areas of interest (*in effect concentrating the wavelet information into the required areas*) prior to the encoding with a version of the progressive SPIHT encoder [6].

Keywords : Target Recognition, Intelligent Compression, Feature detection, Image Processing,

1. INTRODUCTION

The application of this work has been mainly in the area of analysing 8-bit airborne imagery. It is important that we can automatically detect targets of interest and be able to achieve good compression ratios as well as making sure that the important targets and areas of interest are not removed or degraded. A standard quadtree [3] based approach is modified so that it can be used as a target discriminator using the more advanced horizontal-vertical (HV) partitioning scheme. This partitioning attempts to make edges within the image run diagonally through a partition and thus divide a region into two subregions. Results will illustrate how the technique is more flexible than the standard quadtree resulting in significantly less regions in the decomposition. Entropy, approximate entropy and phase congruency have been used as an alternative to quadtrees and can lead to improved region cueing. In particular the phase congruency has proven to be an excellent technique for the detection of fine scale linear features. All techniques have their advantages and disadvantages for this reason a belief network was used to combine results from all three probabilistically.

The cued regions of interest obtained from the feature detection stage are subsequently compressed using an 'intelligent' image compression technique. The technique allows for selective compression to preserve detail in the key areas within the imagery using a wavelet based approach over the entire image but with different levels of compression at different image locations, we are hence cueing the wavelet information into desired areas.

2. FEATURE DETECTION - QUADTREES

Quadtrees are used as an attempt to do a generic and automatic cueing of areas of interest without the need to invent a new set of feature detection algorithms. A quadtree essentially a recursive decomposition of an image in which at each level the image is decomposed in its simplest form by a factor of 2 in both its x and y directions until some suitable criteria is satisfied. Here a textural measure based on edge data and statistical variance within an image portion is used.

More advanced partitioning methods are available such as HV partitioning which is more flexible than the standard quadtree as the position of the partition is not fixed where the image can be partitioned either horizontally or vertically and not by a direct subdivision by 2. The HV decomposition technique attempts to make image edges run diagonally through a partition. In addition there is also triangular partitioning in which the triangles can have any orientation. In the example shown in the left of figure 1 is the image with the HV quadtree overlayed. Results have suggested that the HV technique can potentially offer a decomposition with significantly less rectangles. The HV version here is a modified one in which the algorithm will revert to the standard quadtree if no horizontal/vertical division can be easily detected.

[‡] Defence Evaluation & Research Agency, St Andrews Road, Malvern, Worcs, WR14 3PS, UK.
Email : ducksbury@signal.dera.gov.uk

Wherever possible an image is split into two regions (*either horizontal or vertical*) using an analysis of the two profiles obtained by forming a summation of the current regions gradient information in the appropriate direction. The most dominant profile will ultimately determine the type of decomposition. Initial results show that the technique is feasible. The technique can provide a good approximate cueing for a region of interest.

3. FEATURE DETECTION - ENTROPY

Entropy is a measure of randomness (*or uncertainty*) within some given dataset, the more random the data the more entropy [4,5]. Entropy is defined as

$$H(X) \equiv \sum_{x \in W} p_x \log p_x$$

For example a region with a single gray level value would have a distribution p with just a single peak and an entropy measure of zero. For a section of sky the gray level distribution would be unimodal and therefore have a low entropy measure, whereas for a section with a more typical and varied image content the distribution would be more widespread and have greater entropy. The initial intention was to use the basic entropy as a measure of saliency within an image to highlight areas of interest. Unfortunately in the experiments carried out the standard entropy was very good at excluding areas (*e.g.* sky) but had a tendency to include too much other information within an image and any subsequent mask generated from this would be swamped by potential areas of interest.

Approximate Entropy [4,5] is a measure of unpredictability of a sequence of values.

$$f_m = -\frac{1}{N_m} \sum_{i=0}^{N_m} \log C_m^i \quad \text{and} \quad ApEn = f_{m+1} - f_m$$

Where N_m is the number of strings of length m which can be taken from a sequence and C_m^i is the number of strings which match string i .

Approximate entropy is similar to the standard entropy measure except that it is based on the frequencies of spatial configurations of pixels, rather than simply on the relative frequencies of individual pixel gray levels.

Figure 1 (*center*) shows the result of the approximate entropy. The darker areas indicate salient regions within the image and clearly show the main target of interest together with a certain (*but not excessive*) amount of background features.

4. FEATURE DETECTION - PHASE CONGRUENCY

Until recently, feature detection has largely comprised edge and corner detection. Edges are usually considered to be intensity changes that are locally one-dimensional whilst corners are intensity changes that are locally two-dimensional. The designs of most edge detectors have essentially been optimised for the detection and localisation of step changes in intensity. It is however well known that phase carries (*most of*) the important information about a signal [16], most current edge detectors developed for image analysis retain only amplitude information. Kovessi [12] points out, this not only causes them to fail to detect features of interest, it reduces markedly their invariance to intensity contrast, and makes them sensitive to thresholds.

Many intensity changes of interest in our applications are not at all step-like. There is a single characteristic that (*all*) image features (*luminance profiles that humans perceive as places of interest*) have in common, namely that in the (*short term, or windowed* [11]) Fourier domain (*all*) frequency components over a wide range of octaves are maximally in phase [17]. The actual angle at which this phase-congruency occurs is characteristic of the type of feature. eg for a +ve step in intensity all local (*windowed*) Fourier components will have a phase of 0, a -ve step will have phase π , a +ve ridge $\pi/2$, a -ve ridge $3\pi/2$. The observation suggests that the concept of a 'feature' is defined as an image location at which there is a local congruence of phase. The technique is invariant to local image contrast. That is, will generate the same features for compression irrespective of the contrast and brightness of the image. Phase Congruency [17] can be defined as

$$PC(x) = \max_{f(x) \in [0, 2\pi]} \frac{\sum_n A_n \cos(f_n(x) - \bar{f}(x))}{\sum_n A_n}$$

where A_n is the amplitude of the n^{th} Fourier component and $f_n(x)$ is the local phase of the Fourier component at x . The value $\bar{f}(x)$ that maximises the equation is the amplitude weighted mean local phase angle of all Fourier terms at the point being considered. However an alternative [18] is to search for peaks in the local energy function

$$E(x) = \sqrt{F^2(x) + H^2(x)}$$

where $F(x)$ is the signal with the DC component removed and $H(x)$ is the Hilbert transform of $F(x)$. For two dimensional imagery the one dimensional process is computed at a set of different orientations, refer Kovess [12] for full details.

Figure 1 (right) illustrates this technique, standard edge detection techniques such as Sobel and Canny will perform very poorly on fine scale structure present in the images.

5. FUSION - BELIEF NETWORK

Belief networks have been used in image processing for the detection of urban regions, driveable regions in autonomous land vehicle imagery and improved vehicle detection in airborne imagery [13-15]. Pearl's Bayes networks are directed acyclic graphs. In these graphs nodes represent different hypotheses such as the 'belief' in detecting an urban region, a road network, a vehicle etc. A graph G is a pair of sets (V, A) for which V is non-empty. The elements of V are vertices (*nodes*) and the elements of A are pairs (x, y) called arcs (links) with $x \in V$ and $y \in V$. A typical belief equation can be written to obtain the belief at node A based on the observations at B , C and E as follows :

$$BEL(A_i) = a \cdot l_F(A_i) \cdot \sum_{j,k,l} P(A_i | B_j, C_k, E_l) \cdot p_A(B_j) \cdot p_A(C_k) \cdot p_A(E_l)$$

In other words the belief is the resultant product of causal support information, diagnostic support information and prior knowledge. A full derivation of the equations can be found in [2,13,15].

The idea behind using the belief network approach has been to combine multiple sources of information (*evidence*) in a network to achieve a more mathematically sound result than could be achieved by just applying a logical operation to merge all the information (*evidence*) together. Here, two scales proved sufficient together with three states (*low, medium and high*) for the evidence variable that represented the required features, the actual network used being a simple causal tree. Figure 4 shows an outline around a fused region of interest.

6. INTELLIGENT IMAGE COMPRESSION

The motivation for using quadrees, entropy and phase congruency was an attempt to do a generic and automatic cueing of areas of interest without the need to invent new feature detection algorithms. In [1] a number of standard compression algorithms were considered in which the wavelet approach was chosen as the most suitable as we can target the wavelets to key areas and therefore apply different compression ratios in different image. The standard wavelet decomposition [10] is used together with the simplest (*and most localised*) Daubechies-4 filter kernel [9].

Figure 3, shows a standard result from wavelet compression, namely that a large percentage of the wavelet coefficients in the wavelet scale space image can actually be removed yet still produce a visually good reconstruction. It is apparent therefore that a relatively small percentage of the wavelet coefficients are concentrated in the highly salient areas of the image. This is itself the main principle behind the technique used here, i.e. use a target mask generated by some saliency operator to prevent removal of coefficients in key areas of the image. When this is combined with a suitable encoder an efficient progressive target based compression system is produced.

7. ENCODER

The SPIHT (Set Partitioning In Hierarchical Trees) algorithm is a wavelet based compression scheme offering good image quality with high PSNR¹ it is designed for progressive transmission producing a fully embedded coded file. SPIHT [6,7] is based upon an extension to the embedded zerotree wavelet (EZW) work of Shapiro [8]. It is a coding method so any artifacts produced may be due to the wavelet transform and not the coding process. The process differs from conventional wavelet compression in the way in which it encodes the wavelet coefficients. It can also be used for lossless compression as it codes the bits of the image wavelet transform coefficients in a bit-plane sequence, thus if all

¹ $psnr = 20 \log_{10} \left(\frac{2^b - 1}{rmse} \right)$ PSNR is the Peak Signal to Noise ratio and b is the bit depth with $rmse$ being the root mean square error.

bits are encoded the image can be recovered perfectly². As with other techniques error protection is important, but SPIHT's embedded coding scheme means that information is sorted according to importance hence the need for error protection reduces towards the end of the file. The technique is based on three ideas : a tree based representation of the wavelet coefficients, partial magnitude (*so that the most important information may be transmitted first*) ordering of the coefficients and an ordered bit plane transmission of refinement bits for the coefficients. As a result of this, both the encoder and decoder know the structure of the data and the algorithm therefore no explicit ordering information needs transmitting. Hence if a coefficient at a given level in the hierarchy is insignificant then those children of that node are also likely to be insignificant. The algorithm can also be efficiently implemented using a few list structures.

For comparison figure 5 shows a graph of psnr for this approach verses other standard techniques and figure 6 shows results of this with the standard JPEG at a ratio of 50:1 (0.16 bits per pixel) with the psnr for these images being 26.91 and 22.69 respectively.

9. CONCLUSIONS

The improved mask generated by fusing feature detectors and introducing a fuzzy representation now more accurately represents features such as for example the gun barrel, which was partially cropped when using the quadtree generated mask in isolation. The simple fused mask appears better this is due to the additional information present in the image and therefore additional information that requires encoding and transmission. All sets of results produced reconstructed imagery for which it would have been difficult for the human observer to notice any differences.

There are a number of issues worth mention in particular, although object detection is a difficult task it is still worthwhile to produce object masks. If we remove a certain amount of information (*e.g. wavelet coefficients*) then it is more intuitive to remove them from undesirable and less salient parts of an image rather than removing them indiscriminately³. Despite the fact that the use of a state-of-the-art codec is being employed it is still advantageous to attempt to remove non essential information prior to the encoding process. Typically in the examples shown here, to use a SPIHT codec directly without any masks and removal of coefficients would have required an additional ~ 500 Kbits to ~ 1 Mbits. Technique have been shown here that can locate salient parts of an image using several approaches, combine these into a cohesive mask and then use this to cue the wavelet information to obtain improved compression results. Protecting key areas and thus enabling these to be restored more quickly does provide added value to the recognition process. A more detailed version of this paper can be found in [2].

9. ACKNOWLEDGEMENTS

This work was carried out as part of the Technology Group (TG10) of the MoD Corporate Research Programme. The author would also like to thank Vicky Mortimer at Oxford University for supplying a copy of the approximate entropy software.

(c) British Crown Copyright 2001/DERA,

Published with the permission of the controller of Her Britannic Majesty's Stationary Office

10. REFERENCES

1. Ducksbury P.G., 'Target detection and intelligent image compression', SPIE Aerosense 2000, April 24-28, 2000
2. Ducksbury P.G., Varga M.J., 'Feature Detection and Fusion for intelligent Compression', SPIE Aerosense, April 16-20, vol 4380, 200.
3. Samet H., Tamminen M., 'Computing geometric properties of images represented by linear quadtrees', IEEE Trans PAMI, vol. 7, no 2. March 85.
4. Pincus S.M., 'Approximate entropy : statistical properties and applications', Commun. Statistical Theory Meth., 21:3061-3077, 1992.
5. Mortimer V., 'Content-based Image Compression', First year report, Department of Engineering Science, Oxford University, Oct 1999.
6. Said A., Pearlman W.A., 'A new fast and efficient image codec based on set partitioning in hierarchical trees', IEE Trans on Circuits and Systems for video technology, Vol. 6, June 1996, <http://ipl.rpi.edu/SPIHT/spiht0.html>
7. Manduca A., Said A., 'Wavelet compression of medical images with set partitioning in hierarchical trees', SPIE Symposium on Medical Imaging, Cambridge, MA, March 96, <http://ipl.rpi.edu/SPIHT/spiht0.html>

² There are issues of precision here, as theoretically lossless compression is only possible if the coefficients are encoded using infinite precision. The remedy to this is to use an integer multiresolution transformation referred to as the S+P transform. This is claimed to be as efficient as most lossless encoders.

³ The work described here mirrors what is going on within JPEG-2000 - in particular the encoding of arbitrary shaped regions of interest using wavelet based compression although it approaches the problem from a different perspective. This part of the standard should be available later this year for evaluation.

8. Shapiro J.M., 'Embedded image coding using zerotrees of wavelets coefficients', IEEE Trans Signal Processing, vol. 41, pp 3445-3462, Dec 93.
9. Daubechies I., 'Wavelets', S.I.A.M., Philadelphia, 1992.
10. Clarke R.J., 'Digital compression of still images and video', Academic Press, 1995.
11. Mallat S., 'A wavelet tour of signal processing', Academic Press, 1998.
12. Kovesi P., 'Image features from Phase Congruency', Journal of computer vision, MIT Press.
<http://www.cs.uwa.edu.au/~pk/Research/research.html>
13. Ducksbury P.G., 'Parallel texture region segmentation using a Pearl Bayes Network', British Machine Vision Conference, BMVC-93, University of Surrey, Guildford, UK, pp187-196, Sept 93.
14. Varga M.J., Ducksbury P.G., 'Infra-Red Thermography : Techniques and Applications', Handbook of Pattern Recognition and Computer Vision (2nd Ed), Eds. Chen, Pau & Wang, World Scientific, pp 891-924, 1999.
15. Pearl J., 1986, 'Fusion, Propagation and structuring in Belief Networks', *AI*, 29, 241-288
16. Morrone M.C., Burr D.C., 'Feature detection in human vision : A phase-dependent energy model', *Proc Royal Society*, London, B, 235:221-245, 1988.
17. Morrone M.C., Owens R.A., 'Feature detection from local energy', *Pattern Recognition Letters*, 6:303-313, 1987.
18. Venkatesh S., Owens R.A., 'An energy feature detection scheme', Int Conf on Image Processing, pp 553-557, 1989, Singapore.

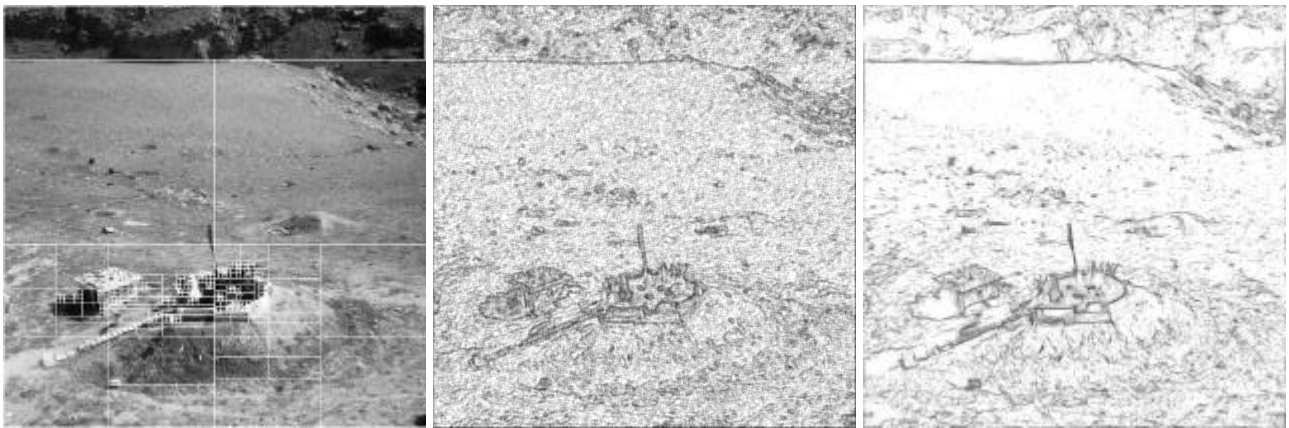


Figure 1; Left - HV quadtree decomposition, Center - Approximate Entropy, Right - Phase Congruency

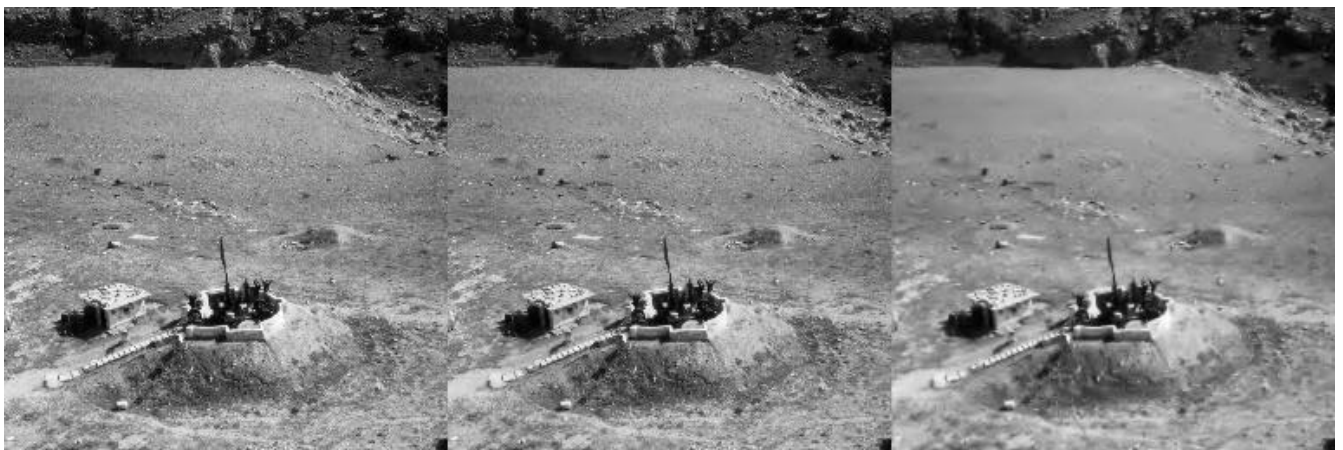


Figure 3, Wavelet reconstruction after removal of 30%, 60% and 95% of weakest coefficients.



Figure 4, Outline of combined mask.

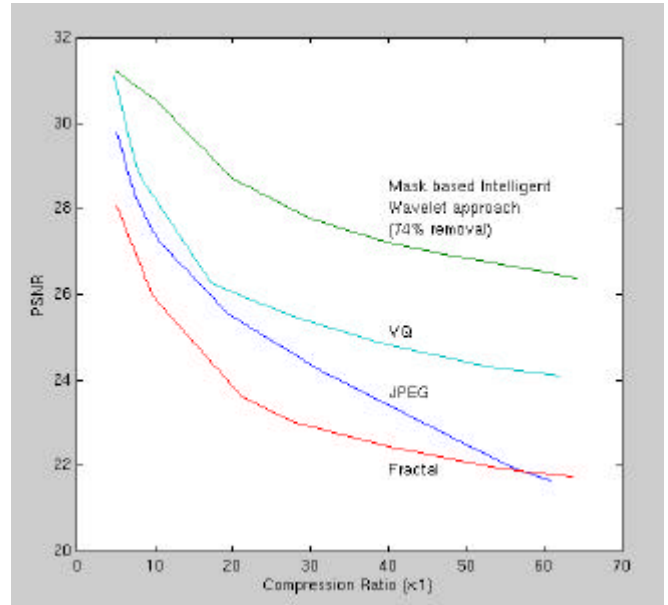


Figure 5, PSNR versus compression ratio



Figure 6, Target Detection and Compression @ 50:1 (0.16 bits/pixel). Left - Combined Fuzzy Mask-based wavelet, Right - JPEG for comparison. The PSNR measurements being 26.91 and 22.69 respectively.