SPIE AeroSense 2000, Vol 4050, 24-28th April 2000, Orlando

Target Detection and Intelligent Image Compression

P. G. Ducksbury

Defence Evaluation & Research Agency, St Andrews Road, Malvern, Worcs, WR14 3PS, UK.

ABSTRACT

This novel approach uses automatic target detection together with compression techniques to achieve intelligent compression by exploiting knowledge of the image content. Two techniques have been experimented with one using horizontal-vertical (HV) partitioned quadtrees the other a varient of entropy called approximate entropy. The object masks that are generated using either of the techniques (*or indeed other feature detectors*) effectively cue potential areas of interest for subsequent encoding using two 'intelligent' image compression techniques. In the first approach, lossless compression algorithms can be applied to regions of interest within the images so that their statistical properties can be preserved to allow detailed analysis or further processing while the remainder of the image can be compressed with lossy algorithms. The degree of lossy compression is dependent both on the information content as well as the bandwidth requirement. In the second approach a wavelet-based decomposition is applied in which selective destruction of wavelet coefficients is performed outside the cued areas of interest (*in effect concentrating the wavelets in required areas*) prior to the encoding with a version of the progressive SPIHT encoder [7,8]. Results will illustrate how both these approaches can be used for the detection and compression of airbourne reconnaissance imagery.

Keywords : Target Recognition, Image Processing, Intelligent Compression, Quadtrees

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. As a result of these requirements coupled with the desire not to invent new and specific feature detectors the use of quadtrees [1] and entropy [2-5] as mediums for locating areas of interest within imagery has been investigated. A standard quadtree [1] 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, if this is not possible our approach will revert to the standard decomposition into four sub-regions of equal size. Results will illustrate how the technique is more flexible than the standard quadtree resulting in significantly less regions in the decomposition. Entropy and approximate entropy has been used as an alternative to quadtrees and as will be shown can lead to improved region cueing.

Various standard compression algorithms where considered including MTWC [6], SPIHT [7,8] which is based upon an extension to the embeded zerotree wavelet (EZW) work of Shapiro [9], CREW [10], Vector Quantisation schemes [11,12], Fractal [13], JPEG, wavelet, as well as the lossless JPEG (JPEG-LS) [14-16]. The cued regions of interest are subsequently encoded using two 'intelligent' image compression techniques.The two options described in this paper allow for selective compression to preserve detail in the key areas within the imagery. The first uses different algorithms at different image locations whilst the second uses a wavelet based approach over the entire image but with different levels of compression at different locations chosen using some saliency operator. In section 2 the target detection using quadtrees is described whilst

Further author information -
Correspondence : P.G.Ducksbury, Email : ducksbury@signal.dera.gov.uk

section 3 describes the use of entropy and approximate entropy. Section 4 discusses the relative merits of the different compression approaches. Section 5 describes the intelligent compression using the detected objects from sections 2 and 3.

2. TARGET DETECTION - QUADTREES

The motivation for using quadtrees [1] was an attempt to do a generic and automatic cueing of areas of interest within both still and video imagery 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 figure 1 an image with the standard and an improved (HV) quadtree overlayed is shown, whilst figure 2 shows part of the actual quadtree[1] for the standard version on the left of figure 1. Results have suggested that the HV technique can potentially offer a decomposition with up 75% 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. 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. In fact it can be seen from figure 1 that the quadtree version on the right is somewhat superior in terms of the reduced number of rectangles required to represent the area of interest.

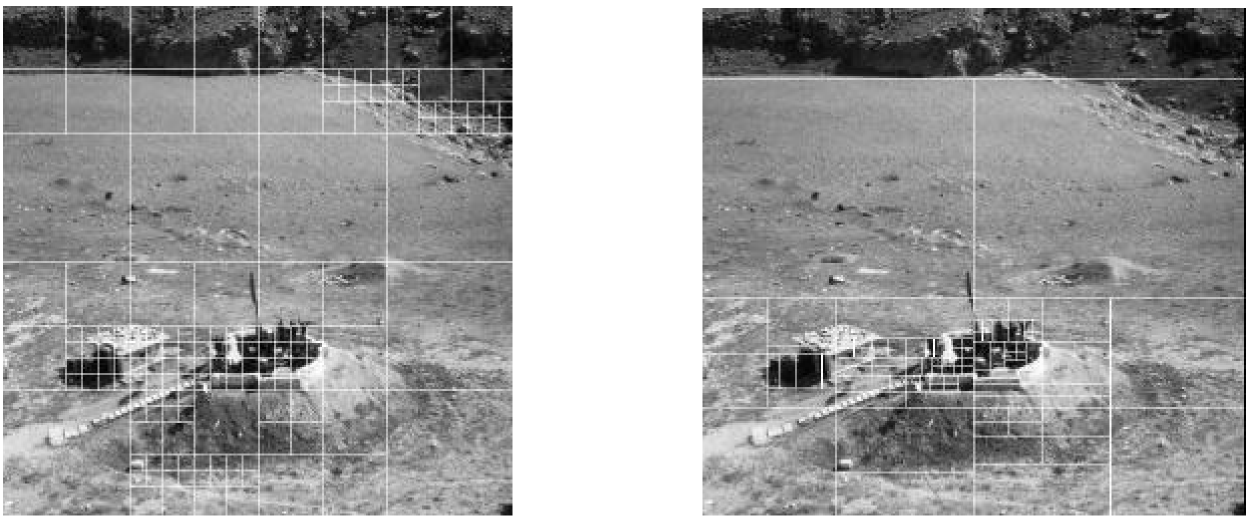

Figure 1, Left - Standard quadtree decomposition, Right - H-V decomposition

1 The tree that is shown here is not fully decomposed to match the image, the nodes indicated in black would be decomposed further. Branches without nodes are leaf nodes with no further decomposition, ordering for the quadtree is top left, top right, bottom left, bottom right.

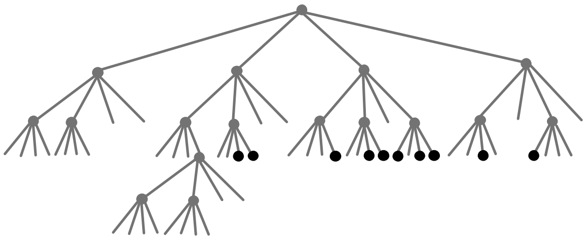

Figure 2, Initial layers of standard quadtree

Figure 3 is another example of both decompositions being applied on some downward looking imagery of a cluster of farm buildings. In this case the H-V version of the algorithm has resulted in approximately 65% less regions.

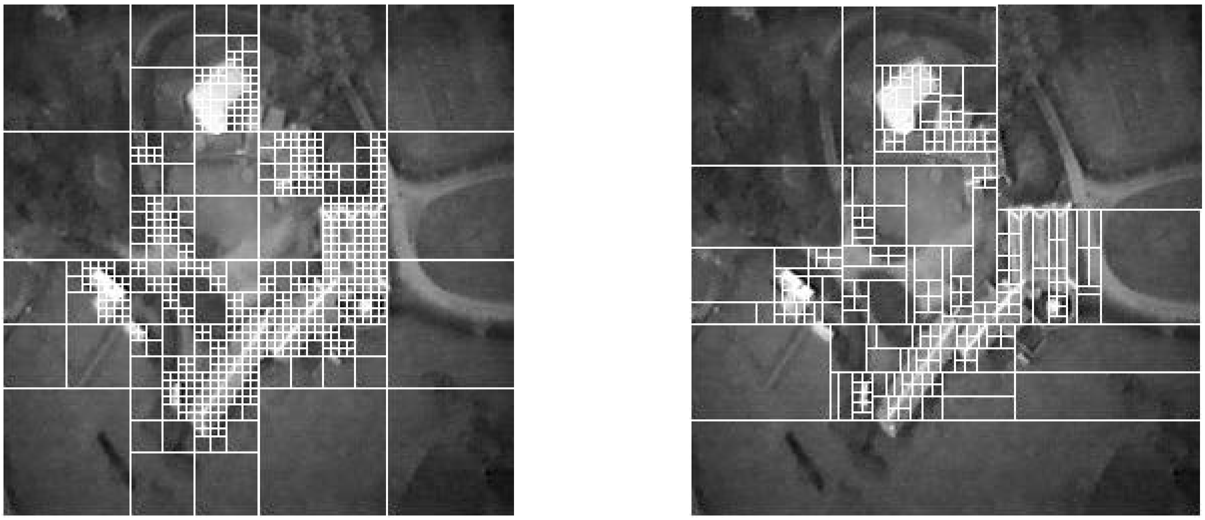

Figure 3, Normal and H-V Quadtree decompositions for potential areas of interest

Another aspect of the use of target detection using quadtrees is shown in figure 4. Here the concept is that if a quadtree is computed on a frame by frame basis from a video sequence then although each tree structure for individual frames is likely to show differences in terms of the number of leaf nodes these differences will not be as significant. We will in effect achieve some invarience to a limited amount of movement within an image[2]. The lower part of figure 4 is an image mosaic [17] made up of approximately 115 frames whilst the upper part shows a plot of the number leaf nodes verses frame number. The two clumps of farm buildings can clearly be seen as they move into the field of view, peak around the center and then move out of the field of view. An adaptive thresholding technique could be used to form the basis of an event detection system.

3. TARGET DETECTION - ENTROPY

Entropy is a measure of randomness (*or uncertainty*) within some dataset, the more random the data the more entropy [2,19]. Entropy is defined as

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

2 For example, consider a tree or bush blowing in the wind, its basic structure does not alter significantly. There will be a transformation from one frame to another but essentially we could approximate this to a spatial shift. The number of leaf nodes will not alter significantly.

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

Approximate Entropy [3,4,5] is a measure of unpredictability of a sequence of values. This is like the standard entropy measure except it is based on frequencies of spatial configurations of pixels, rather than simply on relative frequencies of individual pixel gray levels. Figure 5 shows the result of applying both entropy and approximate entropy to the image of figure 1. The darker areas indicate salient regions within the image and form the basis of a mask which is obtained by applying a smoothing operation followed by thresholding at the $\mu+\sigma$.

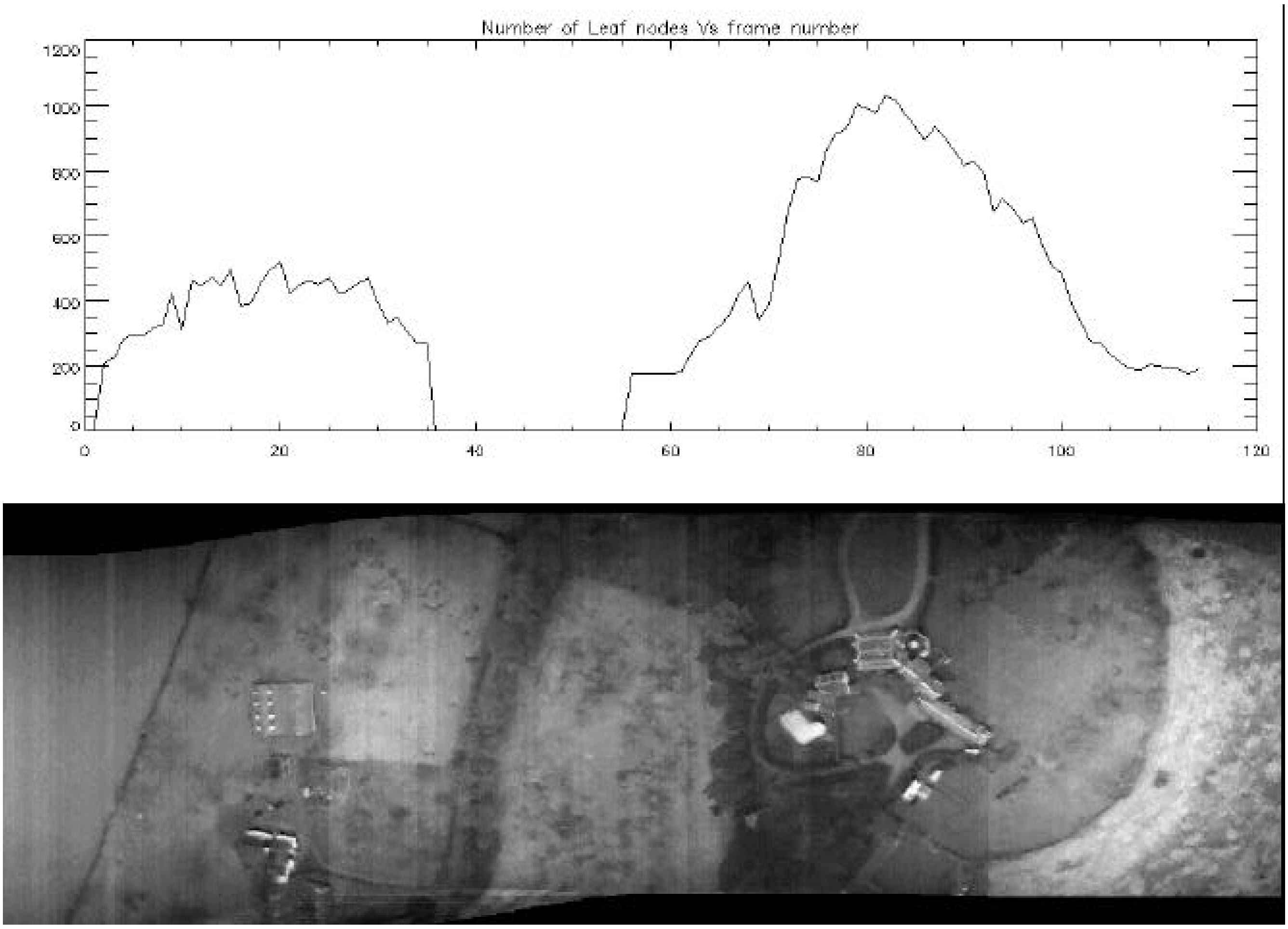

Figure 4, Quadtree based detection in video imagery

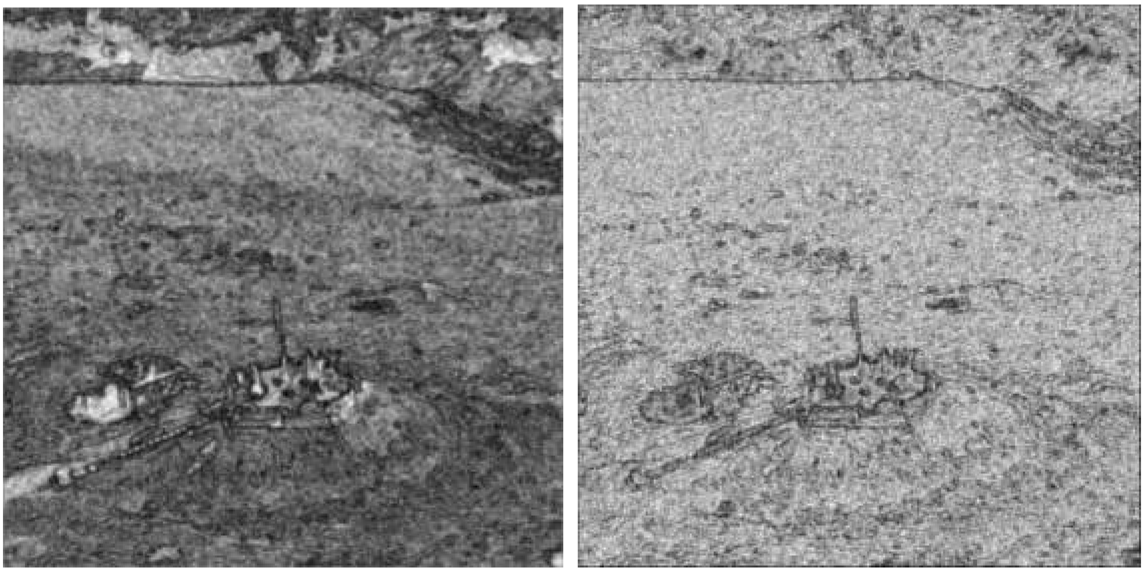

Figure 5, Left – Entropy, Right - Approximate entropy

4. COMPRESSION

A number of standard compression algorithms have been considered including MTWC [6], SPIHT [7,8], embeded zerotree wavelet (EZW) [9], CREW [10], Vector Quantisation schemes [11,12], Fractal [13], JPEG, wavelet, as well as the lossless JPEG (JPEG-LS) [14-16]. For the type of imagery of interest in this work a number of general observations appear applicable. For lossy compression and on images with a rather bland background, reconstruction using the fractal approach appears to be worse than the JPEG. However fractal compression understandably isn't as bad when applied to an image with more structure (*ie content and texture*). The fractal algorithm also appears to degrade more gracefully than JPEG and has less aliasing and ghosting around high frequency components within an image. When applying JPEG to the limits of its compression the equivalent fractal compression is likely to be far more interpretable[3]. When considering the wavelet based approach to compression this appears to out-perform both the JPEG and fractal. Even at high (100:1) compression ratios the results (*although they do contain noticeable artifacts*) are far more readily interpretable. The VQ also proved to outperform both JPEG and fractal and was comparable to the wavelet approach. Indeed at ratios of around 15:1 it proved hard to tell the difference between the two techniques. At 30:1 compression the VQ came out slightly ahead having less artefacts such as ghosting around structures. Both approaches when pushed to the limit could achieve ~100:1 compression with objects of interest still being interpretable by the human visual system. So far for lossless compression, consideration has only been given to GIF and JPEG-LS with the latter achieving ratios of ~ 1.4:1 whilst GIF only achieved 1.1:1 compression[4]. Figure 6 illustrates a comparison of some of these techniques.

5. INTELLIGENT IMAGE COMPRESSION

The motivation for using quadtrees and entropy was an attempt to do a generic and automatic cueing of areas of interest within both still and video imagery without the need to invent new feature detection algorithms. Two approaches have been applied the first one being the application of a lossless based techniques on fine parts of the mesh and lossy techniques on the remainder. Currently the compression algorithms that we have experimented with in the above process have been JPEG and VQ for the lossy part and GIF and JPEG-LS for the lossless. However, on the image in figure 1, compression ratios of about 6:1 are the current best that is achieved using the standard quadtree algorithm and between 11:1 and 14:1 with the improved H-V version. It is obviously the lossless part of the process that is utilising the bandwidth and having the detrimental effect on the compression ratio. Figure 7 shows a reconstruction in which the background has been heavily compressed with JPEG (hence the noticeable artifacts such as the blockiness) whilst the object has been preserved.

---

[3] Interpretability here is a very subjective measure and is taken to mean that the objects of interest in the image can be identified by a human observer at normal distances from a workstation monitor and without zooming the imagery.

[4] This was surprising given GIF's dictionary based method of compression applied on a somewhat bland and low texture background.

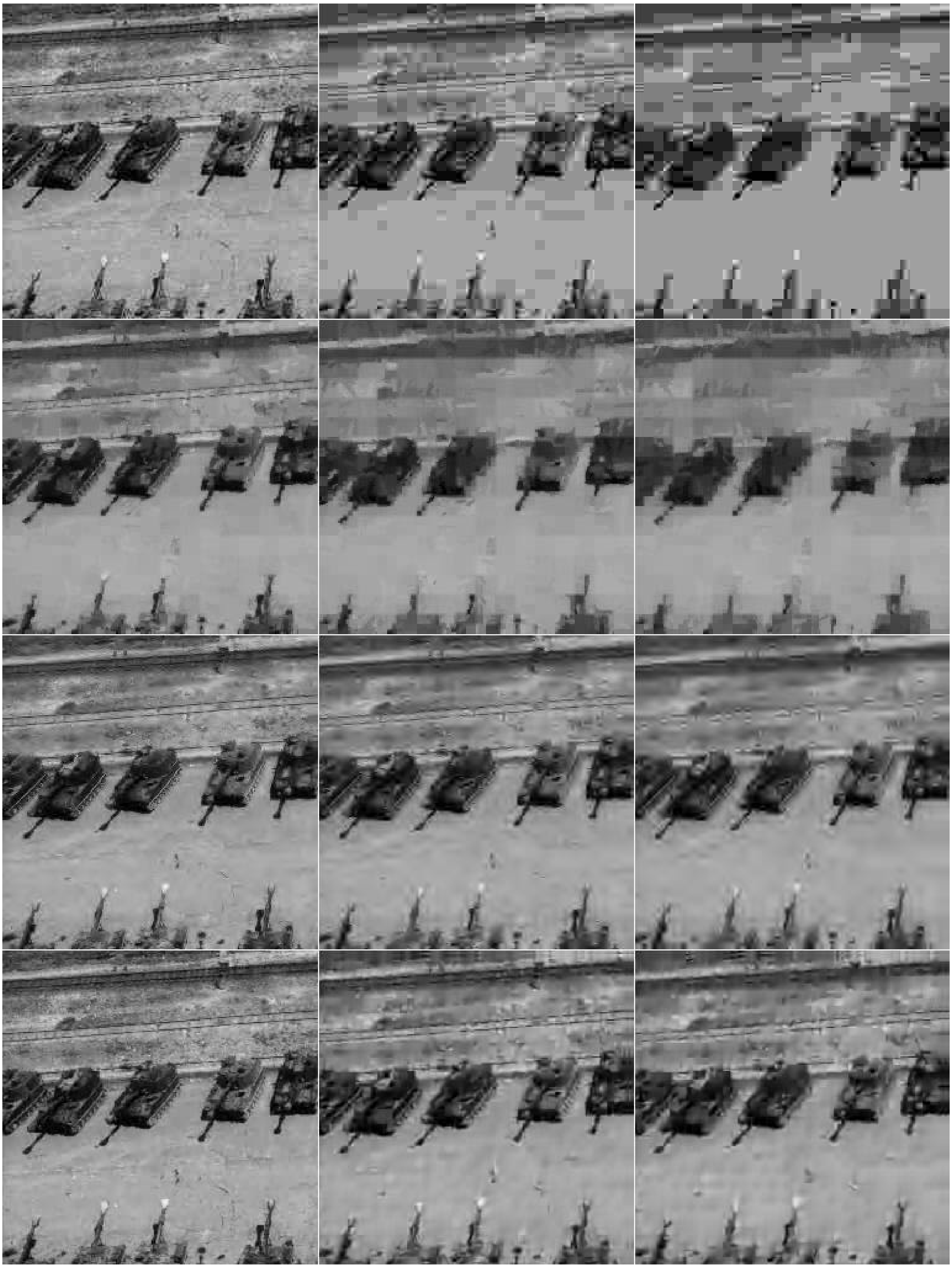

Figure 6. Left column ~10:1, Center column ~30:1, Right column ~50:1.
Top row - JPEG, Second row - Fractal, Third row - Wavelet, Fourth row - Vector Quantisation.

Figure 7, Reconstruction; left 11:1, right 14:1

The second and what is considered the more interesting approach considered here was to use the same compression technique but to effectively apply different compression ratios in different parts of the image. For this the wavelet based decomposition of an image appeared to be the natural choice. The standard wavelet decomposition [21] is used together with the simplest Daubechies-4 filter kernal [18]. Figure 8 illustrates this for a 3 level decomposition in which LH indicates Low-horizontal and High-vertical,HL : High-horizontal and Low-vertical, HH : High-horizontal and High-vertical whilst LL (not shown) is decomposed into the next level of the pyramid. The inverse wavelet transform can then be used to reconstruct the original image.

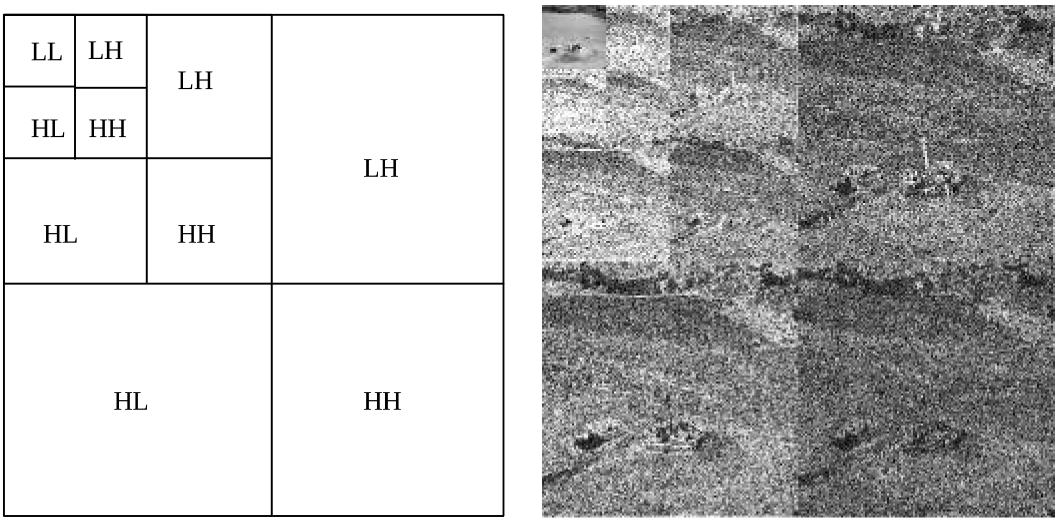

Figure 8, Wavelet Decomposition

Figure 9, shows 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, ie 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 such as the SPIHT algorithm an efficent progressive target based compression system is produced.

Figure 9, Wavelet reconstruction after removal of 30%, 60% and 95% of weakest coefficients.

The SPIHT (Set Partitioning In Heirarchical Trees) algorithm[5] is a wavelet based compression scheme offering good image quality with high PSNR it is designed for progressive transmission producing a fully embeded coded file. SPIHT [7,8] is based upon an extension to the embeded zerotree wavelet (EZW) work of Shapiro [9]. 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 bits are encoded the image can be recovered perfectly[6]. 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[7] ordering of the coefficients and an ordered bit plane transmission of refinement bits for the coefficients.

Figures 10 and 11 show the progressive nature of the SPIHT algorithm for the quadtree and entropy derived masks respectively. For both figures the first column (top - bottom) represents 293:1, 133:1 and 60:1 compression whilst the second column (top - bottom) 40:1, 20:1 and 10:1 compression. We can clearly see improved results in figure 11 by using a more specific mask and somewhat less removal of wavelet coefficients (*in this case 50%*). For example the improved mask now more accurately represents features such as the gun barrel which was partially cropped in the quadtree example. For comparison figure 12 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 25.91 and 24.97 respectively.

6. CONCLUSSIONS

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 indiscriminately. 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. Typically in the examples used here, to use a SPIHT codec directly without any masks

5 The implementation of the SPIHT algorithm used here does not contain any arithmetic coding.

6 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 refered to as the S+P transform. This is claimed to be as efficient as most lossless encoders.

7 ie, the more important information is to be transmitted first.

and removal of coefficients would have required an additional ~ 500 Kbits to ~ 1 Mbits. A technique has been shown that can locate salient parts of an image and then use this to cue the wavelet information to obtain improved compression results.

7. ACKNOWLEDGMENTS

The author would like to thank Vicky Mortimer at Oxford University for running the approximate entropy over the images and Gary Price at DERA Malvern for an initial implementation of the SPIHT algorithm.

8. REFERENCES

1. Samet H., Tamminen M., 'Computing geometric properties of images represented by linear quadtrees', IEEE Trans PAMI, vol 7, no 2. March 85.
2. Gilles S., 'Robust description and marching of images', PhD Thesis, Department of Engineering Science, Oxford University, 1998
3. Pincus S.M., 'Approximate entropy : statistical properties and applications', Commun. Statistical Theory Meth., 21:3061-3077, 1992.
4. Mortimer V., 'Content–based Image Compression', First year report, Department of Engineering Science, Oxford University, Oct 1999.
5. Rukhin A.L., 'Approximate entropy for testing randomness', Tech report, Department of Mathematics & Stastics, Univ of Maryland, 1998.
6. Houng-Jyh Wang et al. 'Multi-Threshold Wavelet Codec (MTWC)', ISO/IEC JTC1/SC29/WG1 N819, Convergence Phase Meeting, Geneva, March 98. http://biron.usc.edu/~houngjyh/cerimon/demo/97-0005/index.htm
7. Said A., Pearlman W.A., 'A new fast and efficient image codec based on set partitioning in heirarchical trees', IEE Trans on Circuits and Systems for video technology, Vol 6, June 1996, http://ipl.rpi.edu/SPIHT/spiht0.html
8. 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
9. Shapiro J.M., 'Embeded image coding using zerotrees of wavelets coefficients', IEEE Trans Signal Processing, vol 41, pp 3445-3462, Dec 93.
10. Boliek M., et al., 'Next generation image compression and manipulation using CREW', ICIP-97, Santa Barbara, CA, Oct 97, http://www.crc.ricoh.com/CREW/
11. Rizvi S.A., Nasrabadi N.M., 'Predictive residual vector quantisation', IEEE Trans Image Processing, vol 4, no 11, pp 1482-1495, Nov 95.
12. Oehler K.L., Cossman P.C., Gray R.M. May J., 'Classification using vector quantisation', In Proc 25th Annual Asilomar Conf on signals and computers, pages 439-445, Pacific Grove, CA, Nov 91
13. Fisher Y., 'Fractal Image Compression', SIGGRAPH 92 course notes, available via http://inls.ucsd.edu/y/Fractals/
14. http://www.hpl.hp.com/loco/ - Lossless JPEG download page
15. Preuss D., 'Comparison of 2-D facsimile coding schemes', IEEE Press, 1976.
16. ISO Final Draft standard, 'Lossless and near-lossless coding of continuous tone still images, JPEG-LS', FCD14495, ISO/IEC JTC1/SC29
17. Kent P. J., 'Multi-resolution Image Registration', IEE Colloquium on Multi-resolution Analysis and Methods in Computer Vision, IEE London, April 1995.
18. Daubechies I., 'Wavelets', S.I.A.M., Philadelphia, 1992.
19. MacKay D., 'A short course on Information Theory', Feb 1995, http://131.111.48.24/pub/mackay/info-theory/course.html
20. Fawcett R., 'Efficient and practical image compression', PhD thesis, Dept of Eng. Sci, Oxford University, Jan 97.
21. Clarke R.J., 'Digital compression of still images and video', Academic Press, 1995.
22. Held G., Data & Image Compression : Tools and Techniques, 4th edition, John Wiley 1996.
23. Schafer R., MPEG-4 : a multimedia compression standard for interactive applications and services, Electronics and Communication Engineering Journal, December 1998., pp 253-262.

24. Rabbani M., Jones P.W., 'Digital Image Compression Techniques', SPIE Tutorial texts in Optical Engineering, Vol TT 7, 1991.
25. Xu X., 'An algorithmic study on lossless image compression', IEEE Data Compression conference DCC-96, Snowbird, Utah, April 96.
26. Mallat S., 'A wavelet tour of signal processing', Academic Press, 1998.

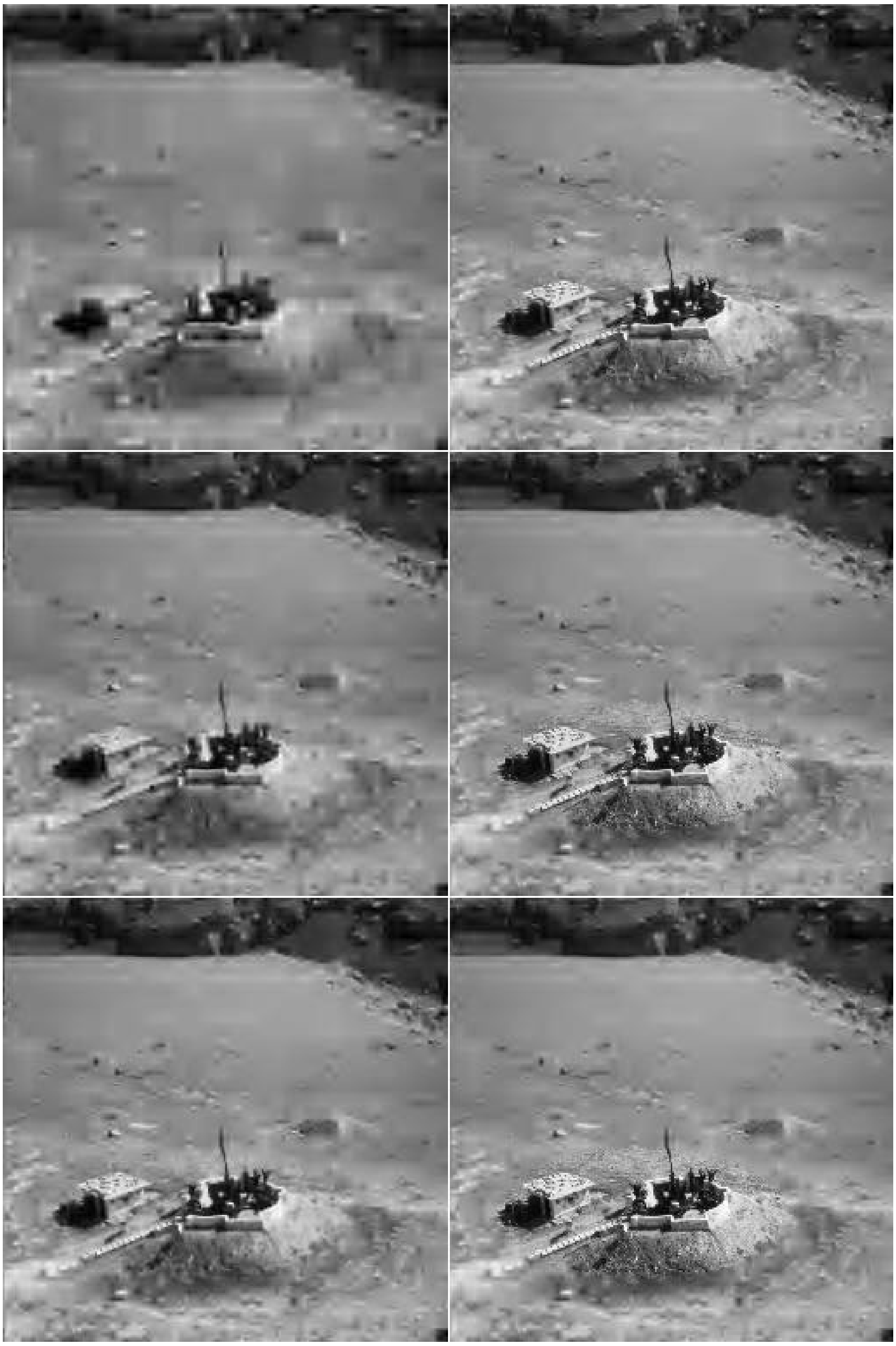

Figure 10, Progressive Target Based Reconstruction - I (*quadtree mask*); Left column, Top – Bottom 293, 132, and 60:1, Right column, Top - Bottom 40, 20 and 10:1 compression.

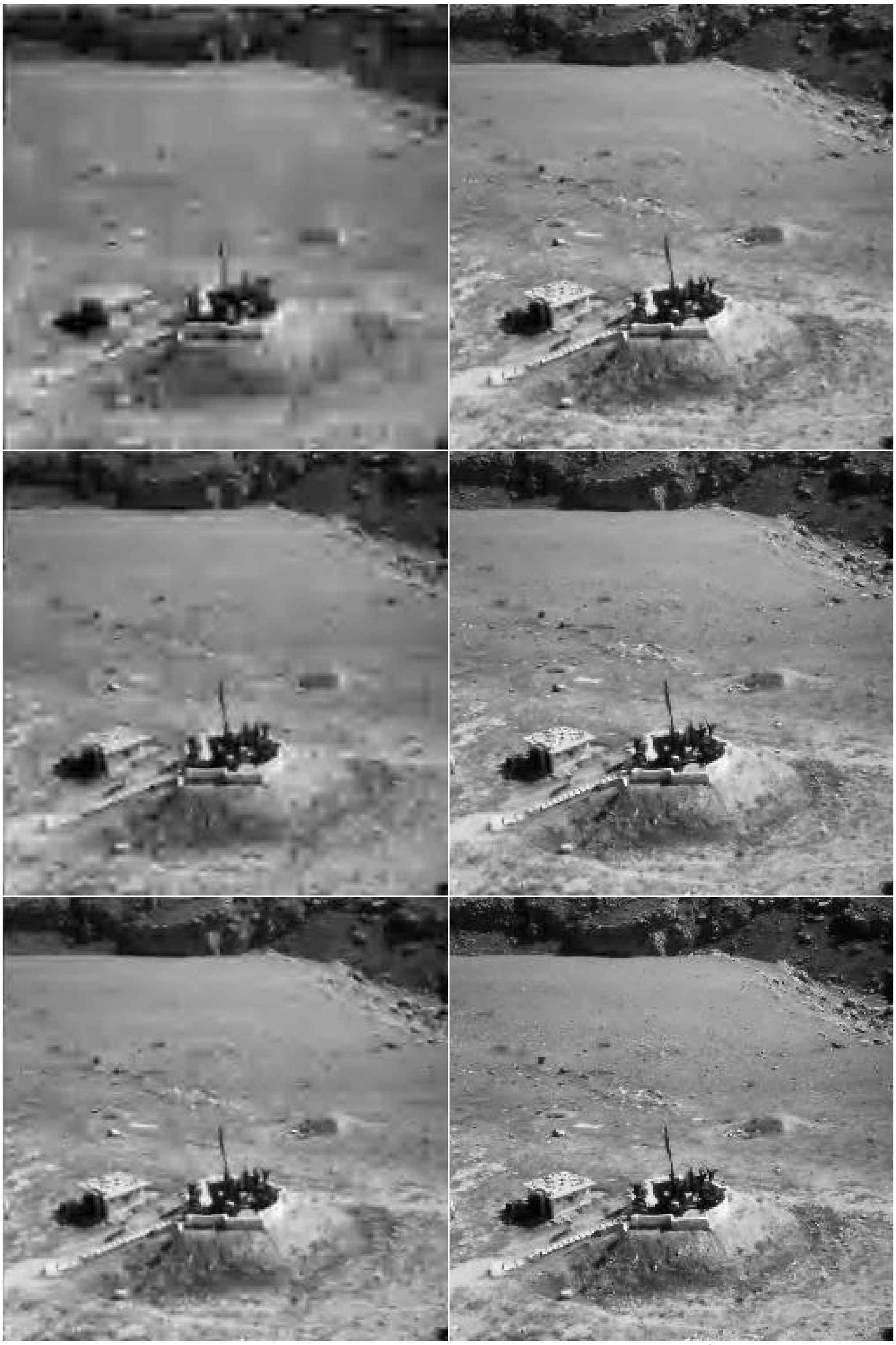

Figure 11, Progressive Target Based Reconstruction - II (*Approximate entropy mask*); Left column, Top – Bottom 293, 132, and 60:1, Right column, Top - Bottom 40, 20 and 10:1 compression.

Figure 12, Target detection and compression @ 50:1 (0.16 bits/pixel). Top - Mask-based wavelet, Bottom - JPEG for comparison. The PSNR measurements being, Top - 25.91, Bottom - 24.97.