Video compression with embedded wavelet coding and singularity maps

Steven Noel[a], Harold Szu[b], Nian-Feng Tzeng[a], Chee-Hung Henry Chu[a], Santipong Tanchatchawal[a]

[a]Center for Advanced Computer Studies, University of Southwestern Louisiana, Lafayette, LA 70504
[b]Systems Research and Technology Department, Naval Surface Warfare Center, Dahlgren, VA 22448

ABSTRACT

We consider the problem of real-time video compression for high-performance wireless information systems, particularly for military applications. Our approach is based on embedded coding of wavelet coefficients. Since this compresses data in order of visual importance, the bitstream can be truncated at any point, allowing a straightforward tradeoff between bitrate and picture quality. For lower bitrates, we propose the inclusion of singularity maps, which are a generalization of edge maps. These singularity maps help identify regions of interest in the video, so that they can be protected while the background is aggressively compressed. To lower costs and increase flexibility, we investigate purely software implementations running on general-purpose processors. Because of the availability of today's fast processors, designed for multimedia and communications applications, such an approach becomes more feasible. To this end we apply integer-to-integer wavelet transforms in order to avoid expensive floating-point operations. We also avoid motion compensation, which requires the expensive computation of motion vectors.

Keywords: Video compression, embedded wavelet coding, integer-to-integer wavelet transforms, singularity maps.

1. INTRODUCTION

Effective video compression is a critical enabling technology for modern multimedia applications. Because of the enormous amounts of data generated by digital video, large compression ratios are required to make most applications feasible. In addition, our society is becoming increasingly dependent on wireless digital communications. While wireless channels offer a number of advantages, particularly high mobility, they often have lower bitrates and are more error prone than wire channels. Real-time video compression for digital wireless channels is particularly critical for a number of military applications.

Our approach to video compression involves embedded coding of the wavelet transform coefficients. Also known as progressive coding, it places bits in order of visual importance, so that lossy compression can be done by a simple truncation of the compressed bitstream. This gives a straightforward way to trade off bitrate and picture quality to best utilize available channel capacity. We choose an encoding scheme based on the EPWIC model of Buccigrossi and Simoncelli.[1] Bitplanes are formed from the quantized wavelet coefficients for video frames, which are then scanned and run-length encoded. Bitplanes are ordered within the bitstream by a greedy algorithm that maximizes the reduction in mean square error per encoded bit. We also propose the use of the reversible TS-transform,[2] a wavelet transform that maps integers to integers, avoiding costly floating-point operations.

For lower bitrates, more aggressive compression is required. Since more time is available per encoded bit, more sophisticated processing can be performed in software. Under these circumstances, we propose the use of singularity maps to improve picture quality for a given bitrate. Such singularity maps are a generalization of edge maps in which edges are represented at multiple resolutions via continuous wavelet transforms. The singularity maps help perform object segmentation, which allows regions of interest to be protected while the background is over-compressed.

Further author information –
S.N. (correspondence): Email: sen7849@cacs.usl.edu; Telephone: 318-482-6147; Fax: 318-482-5791
N.T.: Email: tzeng@cacs.usl.edu; Telephone: 318-482-6147; Fax: 318-482-5791
C.C.: Email: cice@cacs.usl.edu; Telephone: 318-482-6147; Fax: 318-482-5791
S.T.: Email: sxt8393@cacs.usl.edu; Telephone: 318-482-6147; Fax: 318-482-5791
H.S.: Email: hszu@nswc.navy.mil; Telephone: 540-653-2655; Fax: 540-653-2687

Because of the computational demands, real-time video compression has traditionally relied on custom hardware. As Landa *et al* described in detail, the design of such hardware can be a very time consuming, labor intensive, and expensive activity.[3] However, today's fast processors such as Intel's Pentium MMX and Pentium II are designed for better performance in multimedia and communications applications.[4] This makes it more feasible to take a purely software approach, as a cheaper and more flexible alternative to designing custom hardware. The challenge is to develop efficient algorithms that minimize computational complexity while maintaining acceptable quality of the compressed video.

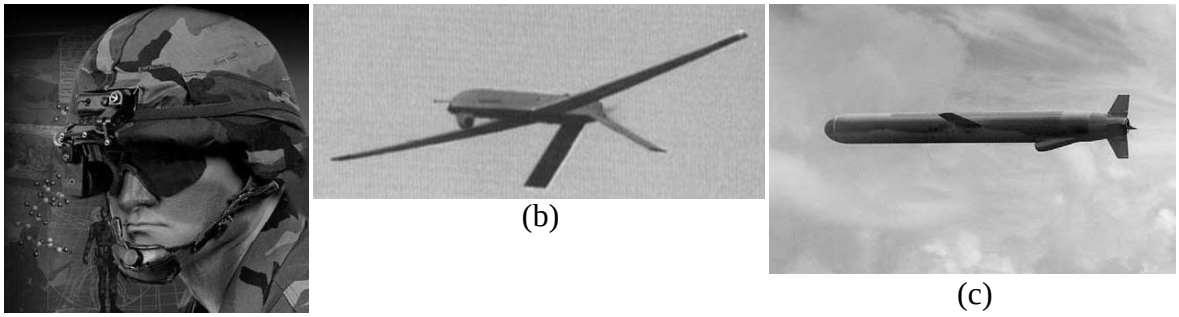
In video compression, for example the MPEG[5] and H.261[6] standards, motion estimation/compensation is often employed to reduce temporal redundancy among video frames. However, the estimation of motion displacement vectors is costly, and can increase coding time by as much as a factor of ten.[7] For this reason we avoid motion compensation for software-only implementations. Indeed, as we show, the bitrate improvement provided by motion compensation can often be small. Also, methods like MPEG employ non-causal bidirectional (forward and backward) prediction, which is far more complex and requires about 6 times more buffer memory than casual (backward) prediction.[7]

In section 2, we describe the need for real-time video compression for wireless military information systems. In Section 3, we describe our wavelet-based video coding/compression scheme. In Section 4, we describe how to enhance this scheme for lower bitrates using singularity maps. In Section 5, we summarize our results and describe plans for future work.

2. MILITARY APPLICATIONS OF DIGITAL WIRELESS VIDEO

We are particularly interested in military applications of real-time video compression over wireless channels. Video and still images are heavily employed in modern high-tech warfare, and wireless links provide a number of advantages, including rapid deployment, high mobility, weather and terrain independence, and flexible network topologies. Moreover, software-only implementations that avoid custom hardware meet the military's policy of using commercial off-the-shelf components to reduce costs.

Some military applications that could benefit from improvements in wireless video are shown in Figure 1. In the near future, even the foot soldier will rely heavily on digital imagery and video. This data will be bidirectional, being transmitted both from and to the battlefield. Unmanned vehicles equipped with a variety of imaging sensors will monitor the battlefield, transmitting images and video to command units. Also, cruise missiles such as the TOMAHAWK currently navigate using stored terrain maps, which limits their use to areas that have been previously surveyed. The capability of transmitting video in real-time would allow next-generation cruise missiles to enter areas that have not been surveyed and be piloted remotely. Wireless video could also benefit a number of civilian applications, such as law enforcement, natural disaster responses, and offshore oil production.



(a) (b) (c)

Figure 1. Example military applications of wireless video: digital battlefield (a), unmanned aerial vehicles (b), and mid-course retargeting for next-generation cruise missiles (c)

We are currently working under a project funded by the Army Research Laboratory to implement wavelet-based video and image compression in software on commercial PCs. Our implementation will follow the IEEE 802.11 standard for wireless local-area networks, using commercially available wireless network PC cards. The Army is likely to accept

the IEEE 802.11 standard, because of its widespread commercial compliance and the Army's commercial off-the-shelf policy. Our developed software can then be used for constructing high-performance wireless information systems.

We also plan to obtain a prototype of WaveNet,[1,8] a software and hardware video processing and compression system that is optimized for wavelet-based operations. WaveNet is being developed for the Army by Trident Systems, Inc. with direct funding from Congress. The rationale behind WaveNet is to provide a general-purpose environment for video applications, thereby reducing the recurring costs of developing custom hardware. This is a similar rationale as for investigating software-only implementations. Indeed, once a system like WaveNet is commercialized, it will become a very attractive option.

3. VIDEO COMPRESSION WITH EMBEDDED WAVELET CODING

We employ an embedded (progressive) wavelet coefficient coding strategy for video compression. This encodes bits in order of visual importance, as measured by mean square error, so that compression can be done by simply truncating the bitstream. Assuming the wavelet coefficients $c_{o,s}(n,m)$ to be represented as 16-bit binary integers, we can write them as

$$c_{o,s}(n,m) = \alpha_{o,s}\sigma_{o,s}(n,m)\sum_k 2^k b_{o,s}(n,m,k).$$

Here $\alpha_{o,s}$ is a scalar multiplier for each subband of orientation *o* and scale *s*, $\sigma_{o,s}(n,m)$ is the sign (± 1), where *n* and *m* index row and column, and $b_{o,s}(n,m,k)$ is the kth bit of coefficient $c_{o,s}(n,m)$.

The wavelet representation for a still image $I(x_1, x_2)$ with spatial dimensions x_1 and x_2 can then be written as

$$\begin{aligned} I(x_1,x_2) &= \sum_{o,s,m,n} c_{o,s}(n,m)\psi_{o,s,n,m}(x_1,x_2) \\ &= \sum_{o,s,m,n} \alpha_{o,s}\sigma_{o,s}(n,m)\sum_k 2^k b_{o,s}(n,m,k)\psi_{o,s,n,m}(x_1,x_2) \\ &= \sum_{o,s,m,n,k} b_{o,s}(n,m,k)\left[\alpha_{o,s}\sigma_{o,s}(n,m)2^k\psi_{o,s,n,m}(x_1,x_2)\right] \\ &= \sum_{o,s,m,n,k} b_{o,s}(n,m,k)\psi'_{o,s,n,m}(x_1,x_2). \end{aligned}$$

Here $\psi_{o,s,n,m}(x_1,x_2)$ are the separable basis functions in the wavelet transform, normalized to have L_2-norm of one. We can interpret this as a representation in which the wavelet coefficients are single bits, where the basis functions are $\psi'_{o,s,n,m}(x_1,x_2)$.

The set of bits at significance level *k* for a particular subband forms a binary image known as a bitplane.[9] In the encoding scheme, bitplanes are sent in order of most to least significance, after being scanned in raster order. A greedy algorithm is used to order the bitplanes across subbands, based on the greatest reduction in mean square error per encoded bit. In other words, each step of the algorithm chooses the subband bitplane producing the most steeply descending rate-distortion curve. The same subband bitplane ordering is used for each frame, until a minimum performance threshold is exceeded (as measured by PSNR), at which point the subband bitplanes are reordered. Once selected for transmission, the scanned bitplanes are run-length encoded. This is done by encoding the length of strings of consecutive blocks of zeros in the bitstream, using block sizes of either 8 or 16 bits.

This encoding scheme is based on EPWIC (Embedded Predictive Wavelet Image Coder) of Buccigrossi and Simoncelli,[1] which has been shown to perform as well as Shapiro's EZW (Embedded Zerotree Wavelet) coder[10] on a variety of test images. The primary difference is that we do not include the non-adaptive arithmetic coding, which uses probabilities calculated from an explicit prior probability model in the wavelet domain. In ordering bitplanes across subbands, EPWIC calculates the number of bits resulting from the arithmetic coding as well as the 8-bit and 16-bit block run-length coding, and chooses the coding with the fewest bits. We found that this was much to slow for a software-only implementation. Arithmetic coding was slower than run-length coding and required the additional computation of a

probability model. Since the performance for run-length encoding was nearly as good, we eliminated the arithmetic encoding.

To reduce costly floating-point operations, we use the RTS transform[2] (reversible TS transform). A reversible wavelet transform is an implementation of an exact-reconstruction transform using only integer arithmetic, allowing a signal with integer coefficients to be recovered without loss. A reversible transform is called "efficient" if the determinant of its transform matrix[11] is approximately ± 1. This means intuitively that a practical entropy coder can efficiently encode its coefficients.

The RTS transform is an efficient reversible version of the TS transform[12] (two-six transform), which is itself an integer version of the (2,6) biorthogonal wavelet transform of Cohen, Daubechies, and Feauveau.[13] Daubechies has recently shown that the TS-transform can be interpreted in terms of wavelet lifting.[14] The RTS transform in one dimension is given by

$$
\begin{aligned}
c_{LP}(0) &= \lfloor (y(0)+y(1))/2 \rfloor \\
c_{LP}(1) &= \lfloor (y(2)+y(3))/2 \rfloor \\
&\vdots \\
c_{HP}(0) &= \lfloor \{-\lfloor (y(0)+y(1))/2 \rfloor + 4(y(2)-y(3)) + \lfloor (y(4)+y(5))/2 \rfloor\}/4 \rfloor \\
&\vdots
\end{aligned}
$$

Here $c_{LP}(t)$ are the low-pass (smooth) wavelet coefficients, $c_{HP}(t)$ are the high-pass (detail) wavelet coefficients, and $y(t)$ is the signal at time t. We apply the standard extension to two dimensions by filtering separately by rows and columns.

Many video coders employ motion estimation and compensation to help reduce temporal redundancy among frames. This generally involves block-matching searches to find motion displacement vectors, which unfortunately can increase the coding time by as much as a factor of ten.[7] Given the current state of technology, motion compensation appears to be infeasible for software-only implementations.

Fortunately, the performance improvements offered by motion compensation can often be fairly small. For example, we compressed the "Claire" videoconferencing video sequence[15] with the PVRG-P64 H.261 coder[16] developed by the Stanford Portable Video Research Group. For each run we varied the motion compensation search diameter from 1 (no motion compensation) to 31 (nearly full search), keeping all other settings at their default. The resulting bitrates are shown in Figure 2(a). While bitrate is minimized for a search parameter value of 3, this is only about a 3% reduction compared to no motion compensation, and only about a 1% reduction compared to full search. Figure 2(b) shows the corresponding PSNR (peak signal-to-noise ratio) for each color component, averaged over all frames. The motion compensation search diameter has essentially no effect on PSNR.

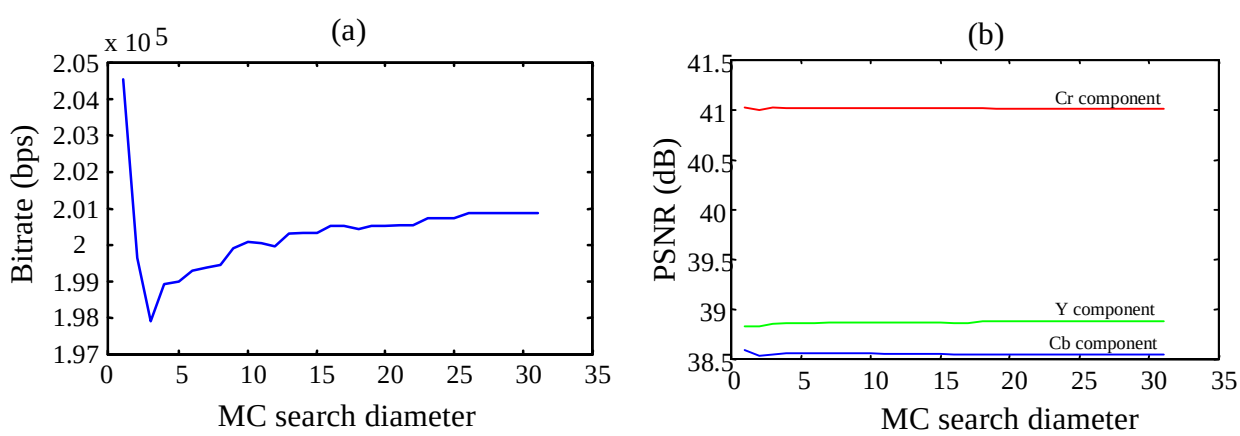


Figure 2. Bitrate (a) and peak signal-to-noise ratio (b) as a function of H.261 motion compensation search diameter

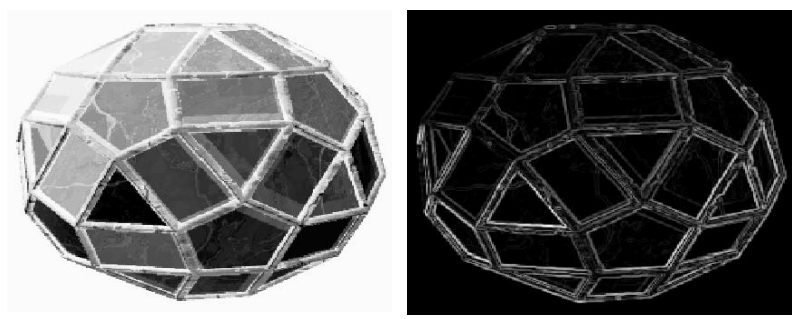
4. VIDEO COMPRESSION WITH SINGULARITY MAPS

At lower bitrates, more aggressive compression of the transmitted video is required, resulting in increased spatial and temporal artifacts. Our strategy is to protect regions of interest at the expense of background regions. We do this by applying a threshold to the image singularity map, a multiscale generalization of an edge map. For military applications, our strategy has the additional benefit of automatic target acquisition. Since it is done at the video sensor, it is a form of distributed target acquisition.

Image segmentation is needed to separate regions of interest from background. The first stage of segmentation is edge detection. A classical approach is based on filters that are smoothed versions of image gradients. An example is the Sobel operator,[17] whose x and y components S_x and S_y are given by

$$S_x = \frac{1}{4}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ and } S_y = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

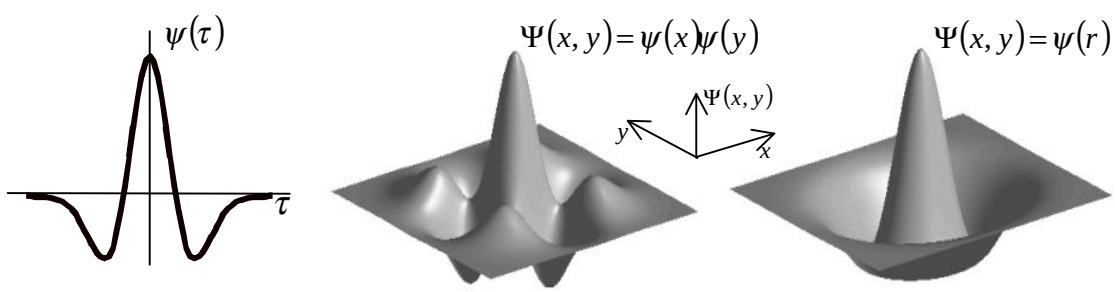
The output is then an appropriate combination of the filter components, such as $\sqrt{S_x^2 + S_y^2}$ or $\max(|S_x|, |S_y|)$.



(a) (b)

Figure 3. Original image (a) and image edges (b) computed with Sobel filter

Mathematically, edges are defined in terms of singularities, which are points at which derivatives are undefined. This is also consistent with the idea of singularities in physics. Edge singularities are difficult to detect in the presence of noise,[18] since the noise itself is filled with singularities. A more robust approach is based on a multiscale generalization of edges that we call singularity maps. Such maps use the wavelet transform to give image edges at a particular scale, as pioneered by Marr.[19]



(a) (b) (c)

Figure 4. Mexican hat wavelets: one-dimensional (a), separable (b), circularly symmetric (c)

The Mexican hat wavelet $\psi(\tau)$ in one dimension is

$$\psi(\tau) = [1 - \tau^2]e^{-\tau^2/2}.$$

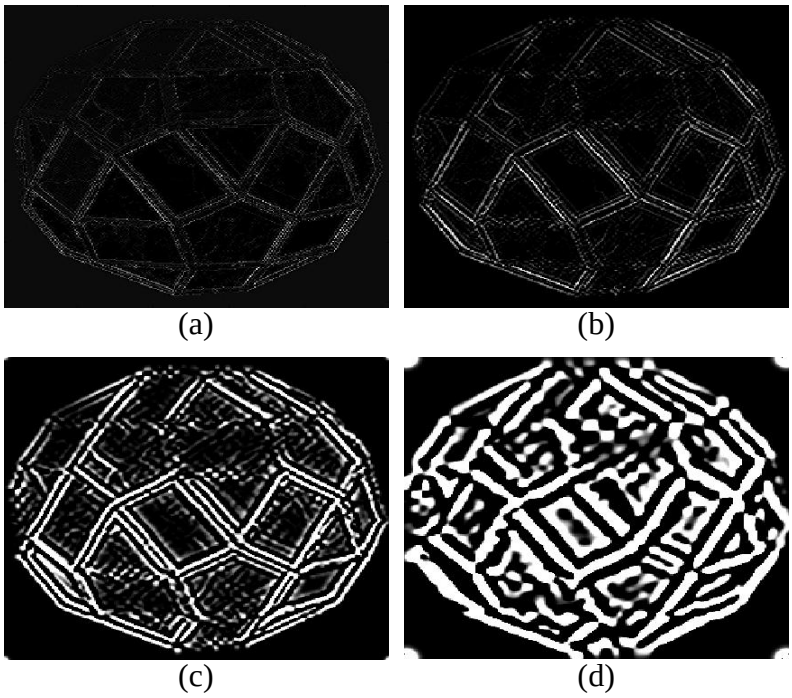


Figure 5. Multiscale singularity detection with separable Mexican hat: wavelet sample points $N = 5$ (a), $N = 9$ (b), $N = 17$ (c), and $N = 33$ (d)

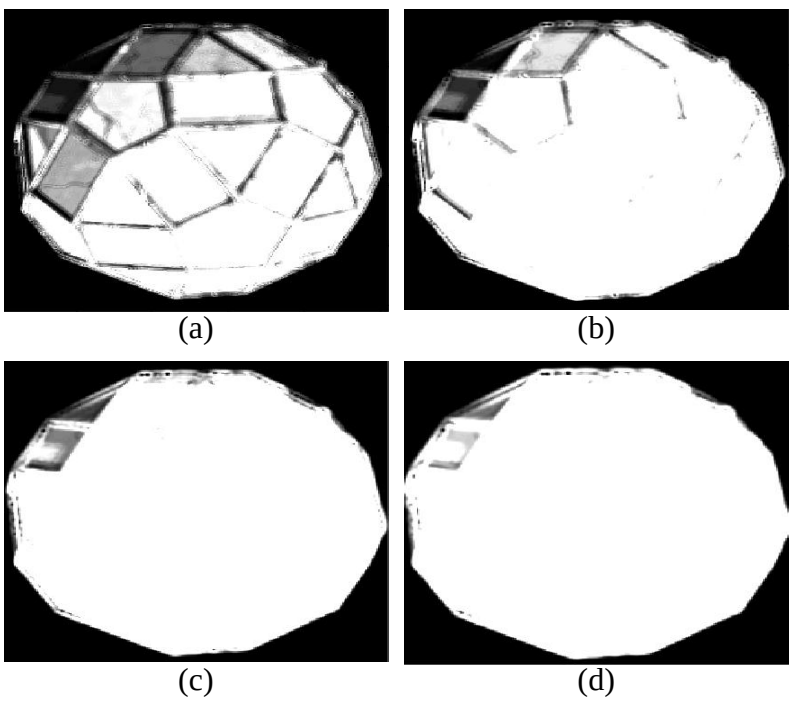


Figure 6. Multiscale singularity detection and contrast enhancement with circular Mexican hat: wavelet sample points $N = 5$ (a), $N = 7$ (b), $N = 9$ (c), and $N = 11$ (d)

One way to extend $\psi(\tau)$ to two dimensions is to construct a separable wavelet $\Psi(x, y)$ via the product

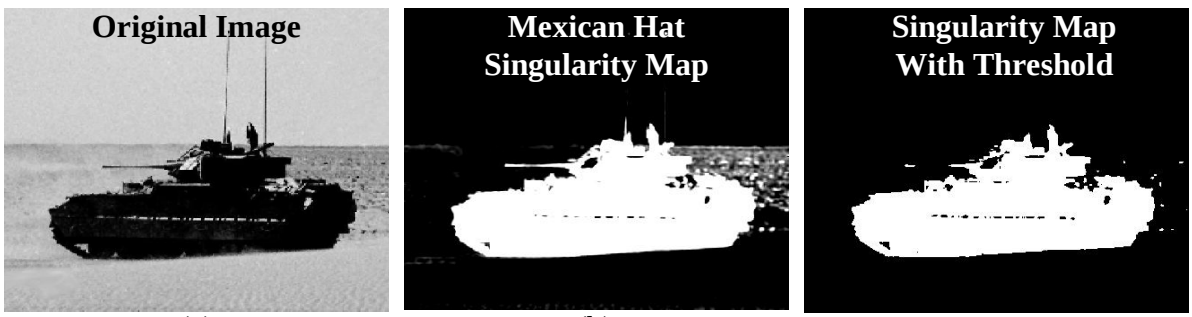
$$\Psi(x, y) = \psi(x)\psi(y).$$

Through the 2-dimensional wavelet transform, this wavelet acts as multiscale singularity detection filter. Wavelet dilation determines the scale of the detected singularities, which for the discrete transform depends on number of wavelet sample points *N*, as shown in Figure 5. Another way to extend $\psi(\tau)$ to two dimensions is to form a circularly symmetric wavelet

$$\Psi(x, y) = \psi(r),$$

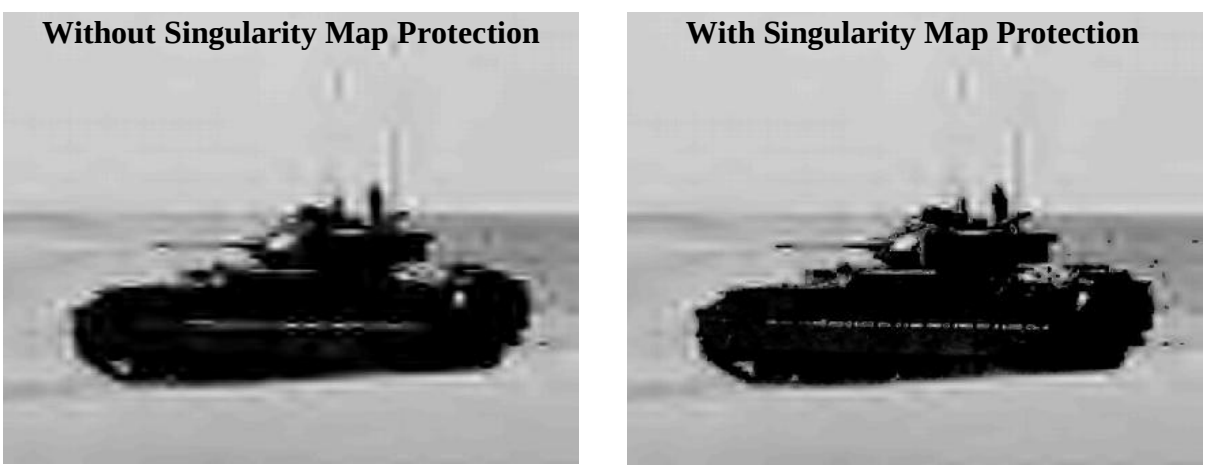
where $r = \sqrt{x^2 + y^2}$. With an appropriate post-filter bias, this wavelet performs scale-dependent contrast enhancement in addition to edge singularity detection, as shown in Figure 6.

A threshold can be applied to the circular wavelet singularity map in order to separate images into regions of interest and background regions, as shown in Figure 7. It could be either a hard step function threshold for a binary separation of classes, or a soft threshold in the sprit of fuzzy class membership. In Figure 8, we show how singularity map thresholding can be applied for region-of-interest protection during compression.



(a) (b) (c)

Figure 7. Separation of region of interest and background via singularity map threshold



(a) (b)

Figure 8. Compression with and without singularity map protection

As Figure 9 illustrates, this simple scheme of wavelet filtering and thresholding rather robust with respect to background noise. However, it is inadequate when the spatial scale and gray or color (e.g. luminance) values of the region of interest are similar to those of the background. In such cases a computer vision approach to region-of-interest determination is appropriate. That is, low-level features that effectively discriminate between region of interest and background could form inputs to neural network classifiers. For example, the features could be wavelet transform coefficients at various scales and orientations. The features could also be moment-based for shape discrimination, or texture-based if the regions-of-interest have distinctive textures.

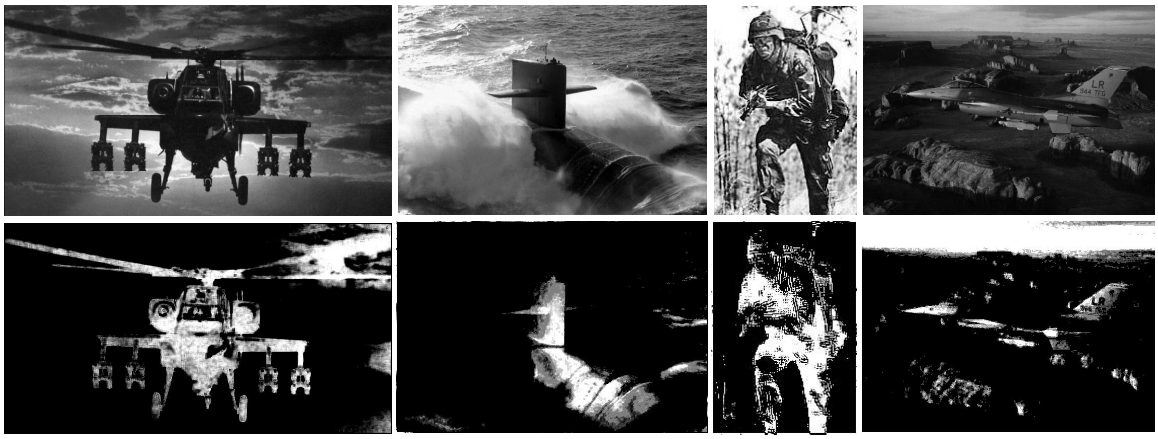


Figure 9. Singularity map region-of-interest determination for a variety of images

5. SUMMARY AND CONCLUSIONS

We have investigated video compression with embedded coding of wavelet coefficients and wavelet-based singularity map protection of regions of interest. Our intended application is real-time video compression for high-performance wireless information systems, especially for the military. Embedded coding compresses data in order of visual importance, so that a target bitrate can be met by simple truncation of the bitstream. Singularity maps are multiscale generalizations of edge maps that help identify regions of interest in the video, allowing them to be protected while the background is more strongly compressed.

Our embedded coding scheme is a simplification Buccigrossi and Simoncelli's EPWIC model, which has been shown to perform favorably against the well-known EZW coder of Shapiro. We omit the explicit prior probability model and non-adaptive arithmetic coding of EPWIC, since block run-length coding was much faster and performed nearly as well. We showed how the separable 2-dimensional Mexican hat wavelet detects edge singularities at multiple resolutions, and how the circular Mexican hat wavelet performs multiscale contrast enhancement in addition to singularity detection. We also showed the application of a threshold to the circular Mexican hat singularity map for region-of-interest protection during compression.

Our approach was successful in improving the image quality of the region of interest at the expense of background image quality. The application of a threshold to the wavelet-generated singularity map was generally robust in separating regions of interest from the background. It also has the advantage of low computational complexity. However, as we showed, this approach can break down when the spatial scale and gray level of a region of interest and the background are sufficiently alike. Under such conditions, we recommend a more sophisticated and computationally complex computer vision approach, in which neural networks determine regions of interest using low-level features such as wavelet coefficients.

We also considered purely software implementations on today's faster general-purpose processors, in order to lower costs and increase flexibility. For this purpose we proposed using wavelet transforms that map integers to integers, avoiding costly floating-point calculations. We also concluded that given the current state of technology, motion estimation is infeasible for software-only implementations. Fortunately, the improvement offered by motion estimation and compensation can often be small, as we showed in a videoconferencing example.

Wireless channels provide a large degree of flexibility for the numerous military applications of video. The software-only approach also meets the military's policy of using commercial off-the-shelf components whenever possible. Wireless video also meets the critical needs of a number of important civilian applications such as law enforcement, natural disaster response, and offshore oil production. Indeed, effective video compression over wireless channels will continue to be critical in our increasingly technology-dependent society.

REFERENCES

1. R. Buccigrossi, E. Simoncelli, *Image Compression via Joint Statistical Characterization in the Wavelet Domain*, GRASP Laboratory Technical Report #414, Univ. of Pennsylvania, 1997.
2. A. Zandi, J. Allen, E. Schwartz, M. Boliek, "CREW: Compression with Reversible Embedded Wavelets," in *Data Compression Conference*, IEEE, 212-221, 1995.
3. J. Landa, C. Hsu, M. Sandford, D. Drum, "WaveNet Compressed Image Restoration Brassboard Design," in *Wavelet Applications IV*, H. Szu (ed.), Proc. SPIE 3078, 529-538, 1997.
4. A. Peleg, S. Wilkie, U. Weiser, "Intel MMX for Multimedia PCs," *Comm. ACM* 40, 25-38, Jan. 1997.
5. ISO/IEC CD 11172: "Coding of Moving Pictures and Associated Audio – For Digital Storage Media at up to 1.5 Mbits/sec," Dec. 1991.
6. CCITT Recommendation H.261: "Video Codec for Audio Visual Services at $p \times 64$ kbits/s," COM XV-R 37-E, 1990.
7. B. Furht, J. Greenburg, R. Westwater, *Motion Estimation Algorithms for Video Compression*, Kluwer Academic Publishers, Boston, 1997.
8. D. Drum, C. Hsu, "Video Compression and Wireless Narrow-Band Communication," in *Wavelet Applications V*, H. Szu (ed.), Proc. SPIE 3391, 228-232, 1998.
9. A. Telkap, *Digital Video Processing*, Prentice-Hall, Upper Saddle River, NJ, 1995.
10. J. Shapiro, "Embedded Image Coding using Zerotrees of Wavelet Coefficients," *IEEE Trans. Sig. Proc*., 41(12), 3445-3462, Dec. 1993.
11. E. Simoncelli, E. Adelson, "Subband Transforms," in *Subband Coding*, Chapter 4, J. Woods (ed.), Kluwer Academic Publishers, Boston, 1990.
12. A. Zandi, M. Boliek, E. Schwartz, M. Gormish, *CREW Lossless/Lossy Medical Image Compression*, Technical Report CRC-TR-9526, RICOH California Research Center, 1995.
13. A. Cohen, I. Daubechies, J. Feauveau, "Biorthogonal Bases of Compactly Supported Wavelets," *Comm. Pure Appl. Math*. 45, 485-560, 1992.
14. I. Daubechies, "Recent Results in Wavelet Applications," in *Wavelet Applications V*, H. Szu (ed.), Proc. SPIE 3391, 2-9, 1998.
15. Claire test video sequence available from the Rensselaer Polytechnic Institute Center for Image Processing Research at ftp://ipl.rpi.edu/pub/image2/sequence/claire/color/.
16. A. Hung, "PVRG-P64 CODEC 1.1," public software release, available from Stanford Portable Video Research Group at ftp://havefun.stanford.edu/pub/p64/.
17. A. Rosenfled, A. Kak, *Digital Picture Processing*, 2nd edition, Academic Press, New York, 1982.
18. S. Noel, Y. Gohel, H. Szu, "Wavelet Detection of Singularities in the Presence of Fractal Noise," in *Wavelet Applications IV*, H. Szu (ed.), Proc. SPIE 3078, 529-538, 1997.
19. D. Marr, "Early Processing of Visual Information," *Philos. Trans. Roy. Soc. London Ser. B*, 275, 483-524, 1976.