University of Strathclyde Glasgow

Strathprints Institutional Repository

Sheikh Akbari, A. and Soraghan, J.J. (2004) *Wavelet-based video codec using human visual system coefficients for 3G mobiles.* In: 12th European Signal Processing Conference, 2004-09-06 - 2004-09-10, Vienna.

WAVELET-BASED VIDEO CODEC USING HUMAN VISUAL SYSTEM COEFFICIENTS FOR 3G MOBILES

Akbar Sheikh Akbari [1], *Pooneh Bagheri Zadeh* [2], *Ednea Cocharan* [2] *and John Soraghan* [1]

[1] Institute of Communications and Signal Processing, Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, U.K., G1 1XW
phone: +44 141 548 2205, fax: +44 141 552 2487, email: akbar@spd.eee.strath.ac.uk

[2] School of Computing & Mathematical Sciences, Glasgow Caledonian University, 70 Cowcaddens Road Glasgow, U.K., G4 0BA
Phone: +44 141 331 3570, fax: +44 141 331 8445, email: pbz@gcal.ac.uk

ABSTRACT

A new wavelet based video codec that uses human visual system coefficients is presented. In INTRA mode of operation, wavelet transform is used to split the input frame into a number of subbands. Human Visual system coefficients are designed for handheld videophone devices and used to regulate the quantization step-size in the pixel quantization of the high frequency subbands' coefficients. The quantized coefficients are coded using quadtree-coding scheme. In the INTER mode of operation, the displaced frame difference is generated and a wavelet transform decorrelates it into a number of subbands. These subbands are coded using adaptive vector quantization scheme. Results indicate a significant improvement in frame quality compared to motion JPEG2000.

1. INTRODUCTION

A new wavelet based video codec is developed. In INTRA frame mode of operation the codec takes a frame of the video sequence and splits it into subbands. Coefficients in each high frequency subband are quantized using a uniform quantization factor divided by their perceptual weights. The perceptual weights based on the HVS properties for quarter common intermediate format (QCIF) image size are derived experimentally. In INTER frame mode of operation the codec employs adaptive vector quantization and overlapped block motion estimation.

The paper is organized as follows. Human visual system coefficients for quarter common intermediate format (QCIF) image size are first calculated. The complete codec scheme is described next. Finally simulation results and a comparative performance with JPEG2000 image coding for INTRA frames are presented

2. CALCULATION OF PERCEPTUAL WEIGHT FACTORS

The perceptual coding model has been designed for a mobile videophone displaying QCIF video. This corresponds to a physical dimension of 1.8×2.2 inches on the workstation monitor, which is used as the videophone display. Therefore, the pixel resolution r, which is measured in pixels per inch, in both the horizontal and vertical dimensions, will be 80 pixels / inch. Let the viewing distance v, which is measured in meters, be assumed to be 0.30 meters. This distance is a good approximation to the natural viewing distance of a human being using a videophone device. The sampling frequency, f_S in pixels per degree is then given by the following equation [3]:

$$f_S = \frac{2\,v\tan(0.5^{\circ})\,r}{0.0254} \qquad (1)$$

The signal is critically down-sampled at Nyquist rate to 0.5 cycle / pixel. Therefore the maximum frequency represented in the signal is:

$$f_{\max} = 0.5\,f_S \qquad (2)$$

Thus the maximum frequency represented in the QCIF image size with the thirty centimetres distance will be 8.246 cycles/deg. The centre radial frequency for each subband is determined by the Euclidean distance of its centre from the origin where subbands are in a square of length 8.246 and the baseband is in the origin. Figure 1 shows the centre radial frequencies for each sub-band of a three level wavelet decomposition.

The mean detection threshold is the smallest change in a colour that is noticeable by the human observer and is employed to determine the perceptual weight factors. It is a function of spatial frequency, orientation, luminance and background colour. In [1], the initial data were measured in the *xy*Y colour space, where *x* and *y* are the C.I.E. chromaticity coordinates and Y is the luminance. This correspondence presents a coding scheme, which compresses the QCIF video using the YIQ colour space. Therefore, the mean detection thresholds for the YIQ space must be determined before the perceptual weights can be calculated. The mean detection thresholds in the *xy*Y space for the centre frequencies of the subbands are first calculated by

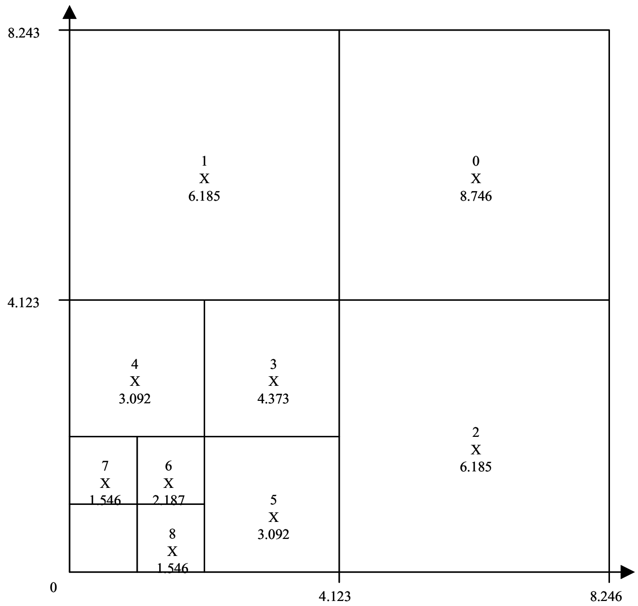

Figure 1: Sub-band centre frequencies in cycles/degree

linearly interpolating the values in [1]. In wavelet decomposition, the diagonal subbands do not discriminate between left and right; so, an average of the two values is employed. The resulting values are first transformed into the XYZ space. The Red-Green line lies approximately in the I direction and the Blue-Yellow line lies mostly in the Q direction. The linear transformations in equations (3) and (4) are used to give two points for each direction in the YIQ space.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.910 & -0.533 & -0.288 \\ -0.985 & 2.000 & -0.028 \\ 0.058 & -0.118 & 0.896 \end{bmatrix} . \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

The YIQ mean detection threshold for each direction is the Euclidean distance between these two points. The computed weighting factors for each subband of QCIF video, based on the properties of the HVS, are shown in Table 1.

3. WAVELET BASED VIDEO CODING USING HVS COEFFICIENTS

The wavelet based video coding scheme is shown in Figure 2. Since subsequent frames are predicted from INTRA- frames, it is important that the quality of the INTRA frame is as high as possible. Hence, the perceptual weights are used in this mode of operation to improve the visual quality of the coded sequence.

In INTRA frame mode of operation, an INTRA frame is input to the codec. The codec performs three levels lossy

DWT LEVEL	SUBBAND	Y-DOMAIN	I-DOMAIN	Q-DOMAIN
ONE	LH	4.3807	2.0482	1.0502
	HL	3.4573	1.6159	1.0992
	HH	1.2372	0.6978	0.6065
TWO	LH	5.9673	3.6449	2.6340
	HL	6.1708	2.7149	2.4728
	HH	4.1934	1.6384	1.1331
THREE	LH	6.1796	4.5685	7.1443
	HL	6.1984	3.3243	5.5495
	HH	5.3931	2.9888	2.2339

Table 1: Perceptual weight factors for the YIQ colour domain

JPEG2000 wavelet transform [4] on the INTRA frame to decorrelate it into ten frequency subbands. The lowest frequency subband is coded using lossless JPEG2000 coding scheme. The high frequencies subbands are pixel quantized using a uniform quantization factor divided by their perceptual weights to regulate the quantization step-size in the quantization of the wavelet coefficients and minimize the visibility of compression artefacts. The perceptual weights used are specific for visual communication using handheld videophone devices. The output of the quantization process is a significant map and some quantized coefficients for each subband. The significant map is a binary image where a '1' corresponds to a pixel that has non zero value and will be used in the next stage and a '0' corresponds to a pixel that has a zero value. For each binary map, the number of zeros and ones are calculated and this map is coded using quadtree structure [5] for the group with the higher population.

In INTER frame mode of operation, the displaced frame difference (DFD) is input to the DWT. The DFD frame is generated from the difference between the current input image and the predictive frame, which is the motion compensated version of the reconstructed previous frame. The codec performs three level lossy JPEG2000 wavelet transform to the DFD frame and decorrelates it to ten frequency subbands. The lower frequency band is coded using lossless JPEG2000 coding scheme. The coefficient of high frequency subbands are first split into vectors and the most significant vectors according to their energy are preserved. A measure of the significance of a vector of DWT coefficients of a subband k is given by the normalized energy of the vector, shown in equation (5):

$$E = \frac{1}{N_k \times N_k} \sum_{i=1}^{N_k} \sum_{j=1}^{N_k} X_k(i,j)^2 \quad (5)$$

where $X_k(i,j)$ are the wavelet coefficients for subband k and $N_k \times N_k$ is the vector size. The vectors with the highest energy measures are preserved for the vector quantization stage. The output of the vector selection process is a significant map for each subband. The significant map is a binary image where a "1" corresponds to a vector that has been preserved and will be used in the VQ stage, while a "0"

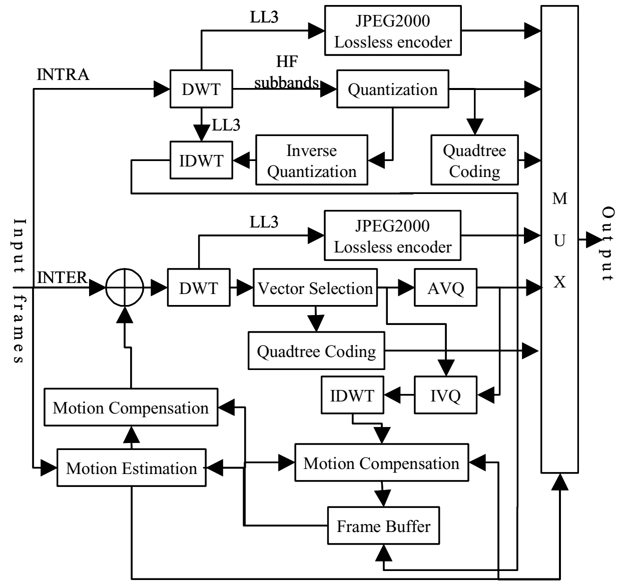

Figure 2: AJVQ encoder

corresponds to a vector that has been discarded. For each binary map, the number of zeros and ones are calculated and this map is coded using quadtree structure [5] for the group with the higher population. The selected vectors are coded using AVQ scheme, which were presented in [2].

4. RESULTS

The Y frames of the five QCIF test sequences, 'Carphone', 'Foreman', 'Miss America', 'mthr_dotr' and 'Suzie' were coded at 32kbits/s using our wavelet based video codec for INTRA and INTER frames. These sequences were also coded at 32kbits/s using JPEG2000 image codec [6] to compress the INTRA frames and the new codec for the INTER frames. The PSNR measurements for the first INTRA frames are given in Table 2. To illustrate the improvement in visual quality obtained when using the perceptual model, the first INTRA frame from the 'Foreman' sequence of both coding scheme is shown in Figure 3 (a) and (b). The picture quality in (b) is much sharper than in (a), with less blurred edges and better facial detail.

REFERENCES

[1] Van Dyck, R.E. and Rajala, S. A.: 'Subband/VQ Coding of Colour Images with Perceptually Optimal Bit Allocation', IEEE Trans. on Circuits and Systems for Video Technology, February 1994, vol. 4, no. 1.
[2] Voukelatos, S.P. and Soraghan, J. J.: 'Very Low Bit Rate Colour Video Coding Using Adaptive Subband Vector Quantization with Dynamic Bit Allocation', IEEE Trans. on Circuits and Systems for Video Technology, April 1997, vol. 7, no. 2, pp. 424-428.
[3] Nadenau, M. J. and Reichel, J.: " Wavelet-Based Colour Image Compression: Exploiting the Contrast Sensitivity Function", IEEE Transaction on Image Processing, January 2003, vol. 12, no. 1.
[4] Lawson, S. and Zhu, J.: 'Image compression using wavelets and JPEG2000: a tutorial', Electronic & Communication Engineering Journal, June 2002,vol. 14, no. 3. pp. 112-121.
[5] Samet, H.: "The quadtree and related hierarchical data structures", ACM Computer Surveys, vol.16, no. 2, June 1984.
[6] Lurawave Smart Compress available for free download from: www.luratech.com.

Figure 3a Figure 3b

Figure 3: Comparison of visual quality

	PSNR (dBs)	
Sequence name	JPEG2000	Our system
Carphone	31.64	34.57
Foreman	32.51	33.30
Miss_am	42.16	44.97
Mthr_dotr	33.28	35.02
Suzie	36.51	37.83

Table 2: PSNR values for coded sequences with JPEG2000 and our coding scheme