Jurnal Teknologi, 34(D) Jun 2001: 41–62
© Universiti Teknologi Malaysia

USING MULTISCALE EDGE DETECTION FOR LOW BIT-RATE SUBBAND VIDEO CODING

MOHSEN ASHOURIAN[1], ZULKALNAIN MOHD. YUSOF[2], SHEIKH HOSSEIN S. SALLEH[3], SYED ABD. RAHMAN A. BAKAR[4]

Abstract. This paper introduces a video compression system for very low bit-rate video-conferencing and tele-monitoring applications. The video codec first performs a three-dimensional subband decomposition on a group of video frames, and then encode the subbands with different quantization methods. By using a cubic spline wavelet, the spatial filter bank acts as a multiscale edge detector. This property is used for efficient selection and coding of high frequency subbands with geometric vector quantization. For lowest tempo-spatial subband, a DPCM coding with an entropy coder was used. Results at several low bit rates (16, 32, 64, 128 Kbps) are reported and compared with H.263, the standard for low bit rate video coding.

1.0 INTRODUCTION

There are different types of video compression systems based on their design and application. This paper explains results on developing a video compression system with low and variable bit-rate using three-dimensional subband coding techniques. In three-dimensional subband coding both temporal and spatial redundancy are reducing using the similar method [1–2]. The major challenge in subband coding is proper selection of filter banks for decorrelating information among subbands and encoding each subband based on its statistical properties. The major contribution of this paper is using an efficient selection and coding of edge information in subband transform domain for compression of high temporal subbands while maintaining their perceptual information.

The rest of this paper is organized as follows. Section 2 discusses the first stage of proposed video coder which is a three dimensional filter bank, section 3 explains different encoding methods used for different subbands, and in section 4, the rate control mechanisms are explained. Section 5 provide a rough approximation of coder complexity and finally section 6 provides the practical results and section 7 concludes the paper.

[1] Faculty of Engineering, Azad University of Iran, Majlesi Branch, Esfahan, Islamic Republic of Iran. E-mail: mohsena@ieee.org

[2,3&4] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor Darul Ta'zim, Malaysia. E-mail: zul@suria.fke.utm.my

2.0 THREE DIMENSIONAL FILTER BANK

In three-dimensional subband filtering the digital video signal is filtered and subsampled in all three dimensions (temporally, horizontally and vertically) to yield the subbands, from which the input signal can be losslessly reconstructed in the absence of coding loss.

The diagram in Figure 1 illustrates the specific filter bank structure chosen for the results presented here, which consists of 11 spatio-temporal frequency bands. This

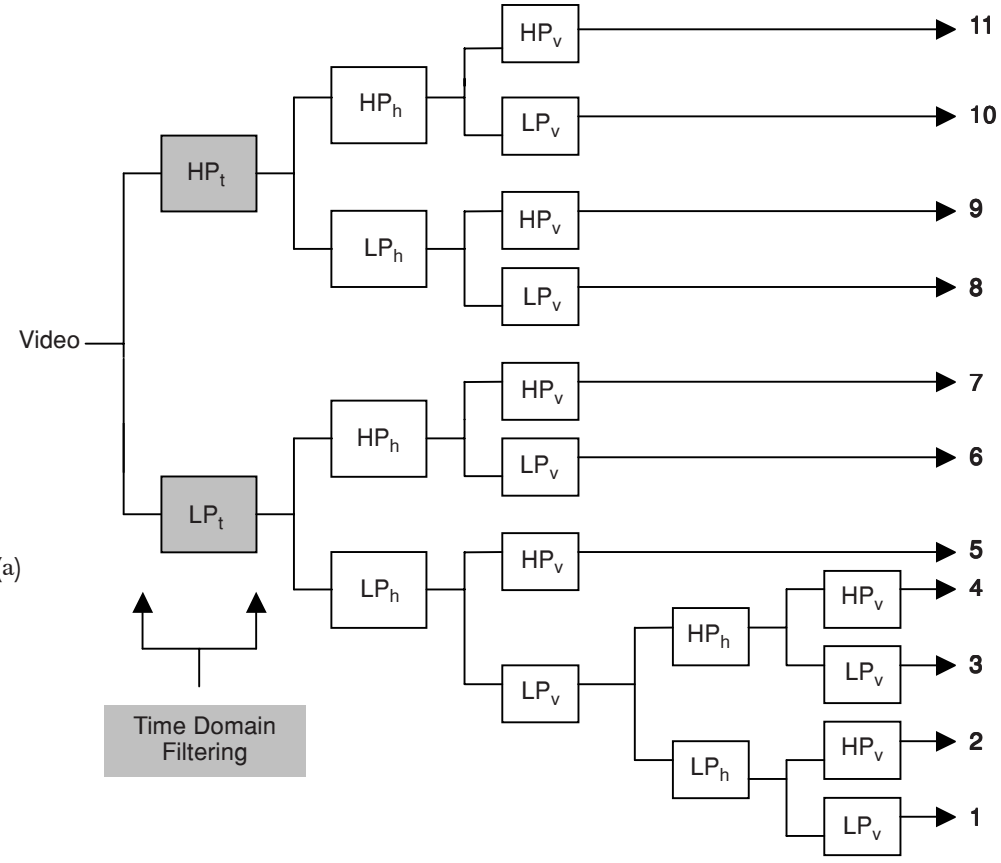


Figure 1 Selected 3-D Filtering ((*a*) Structure, (*b*) Frequency Map)

contains two temporal subbands and several spatial subbands in each temporal band. The terms HP and LP refer to high-pass filtering and low-pass filtering, where the subscripts t, h, and v refer to temporal, horizontal, and vertical filtering respectively. The image frames are filtered temporally using the two-tap Haar basis functions, which are basically the difference and average between frames, producing a high-pass and low-pass temporal frequency band [3]. Temporal decomposition is followed by horizontal spatial filtering and vertical spatial filtering. In proposed system, a cubic spline wavelet coefficient filters is used for spatial filtering in both high temporal and low temporal frequency bands in order to do a multiscale edge detection process over subbands. The basic idea behind using wavelet for multiscale edge detection is choosing the corresponding wavelet (that related to HP and LP filter) to be first derivative of a smooth function [4–6]. Table 1 shows the analysis and synthesis filter related to this wavelet transform [7].

Table 1 Spatial analysis/synthesis filter used in variable rate system

N	LPF Analysis	HPF Analysis	LPF Synthesis	HPF Synthesis
1	0.17677669	–0.35355339	–0.35355339	–0.17677669
2	0.53033008	–1.06066017	1.06066017	0.53033008
3	0.53033008	1.06066017	1.06066017	–0.53033008
4	0.17677669	0.35355339	–0.35355339	0.176776695

3.0 ENCODING OF VIDEO SUBBANDS

The basic idea of subband coding is to decompose the full-band input signals into a number of frequency subbands and then code each subband separately. The following section explain different types of quantization methods used for different subbands.

3.1 Lowest Tempo-Spatial Subband

The lowest tempo-spatial frequency subband (Band 1 in Fig. 1) is very important in image and video coding. Its distribution is highly image dependent and cannot be modeled as Gaussian or Laplacian, so an optimized Lloyd-Max quantizer cannot be use directly for its quantization.

In proposed system, a DPCM coding with addition of an entropy coder is used for compression of this band [2, 8]. The proposed DPCM coder was a seven order predictor, (Equation 2) which use dependencies of pixels in time and space for coding prediction error (Equation 1).

$$d(i,j,t) = x(i,j,t) - x_p(i,j,t) \tag{1}$$

$$\begin{aligned} x_p(i,j,t) &= e_1 x(i,j-1,t) + e_2 x(i-1,j,t) \\ &+ e_3 x(i-1,j-1,t) + e_4 x(i,j,t-1) + \\ &e_5 x(i,j-1,t-1) + e_6 x(i-1,j,t-1) + \\ &e_7 x(i-1,j-1,t-1) \end{aligned} \tag{2}$$

where e_1 to e_7 are prediction coefficients, and could be changed adaptively or be fixed. In implemented system, in order to have low complexity, fixed coefficient as below were used

$$\begin{aligned} e_1 &= e_2 = e_5 = e_6 = 1/2; \\ e_4 &= 1; \\ e_3 &= e_7 = 1/4; \end{aligned} \tag{3}$$

In order to further increase the coding efficiency of this band an entropy coder was added to the output of DPCM coding of system. Table 2 shows the operation of this entropy coder.

Table 2 Entropy Coder for LFS DPCM Coder

Quantizer level	DPCM code	Entropy code
1	0000	11
2	0001	011
3	0010	1000
4	0011	10011
5	0100	101100
6	0101	101110
7	0110	0100110
8	0111	1010
–8	1000	01000
–7	1001	0100111
–6	1010	1011111
–5	1011	101101
–4	1100	010010
–3	1101	10010
–2	1110	0101
–1	1111	00

3.2 High Frequency Subbands

In general coding efficiency of hybrid video coders which uses motion estimation method is better than 3-D subband coding schemes [9]. In a hybrid coder motion estimation-compensation module selects the blocks of frames that has motion and only sends information for refreshing these blocks in its inter-frame mode. This classification of frame blocks and motion estimation-compensation, makes the encoder complex, and its output bit-stream very sensitive to noise, but on the other hand improves the coding efficiency of system. The main approach in the developed coder in this section, to improve coding efficiency of 3-D subband coder, was devising a mechanism for subbands block classification. Since at low bit-rate, human visual system is more sensitive to distortion of edge rather than texture and background in scene [9], selecting blocks of frames that contain more important edges and coding them properly should be on priority in process of coding. In order to achieve this goal, different approaches has been proposed. In [10] an edge detector was used for lowest frequency subband and in some other investigations a motion estimation method used jointly to 3-D subband scheme [11–13].

The general problem in an edge detection scheme is difficulty in distinguishing sharp variation of signal resulted from real objects edge from changes resulting from noise, texture or generally less important variation in image. Since with proper selection of filter bank, subband coding (or wavelet transform) can do a multiscale edge detection, which could provide a simple and efficient mechanism for selecting the important edges of an image by considering their repetition along consecutive scales and their amplitude. This has been the major concept in the developed video coder in this section.

Coding higher frequency subbands is implemented in two stages. At first the blocks of subbands containing important edges are determined. This process is being done by calculating the amplitude of extreme (based on Equation 2) and comparison of extreme generated in the first scale of spatial decomposition (means band 1, 2, 3 and 4) with its second scale (bands 5, 6, 7). After selection, two different methods were followed for encoding information. In the first system (Figure 2) the encoding of low temporal bands (band 2–7) has been done by direct scalar quantization and entropy coding of multiscale edge information (amplitude and position) and reconstructing the signal at the receiver [4–5], which is complex in decoding process. The second approach (Figure 3) is encoding selected blocks in all high frequency bands using geometric vector quantization (GVQ). The motivation for using GVQ is its high efficiency and accuracy in coding edge information that would be explained later. Both systems were compared with a system similar to what developed in [10] and uses Canny edge detector for lowest band (Figure 4). The systems are named with the following abbreviations for further references.

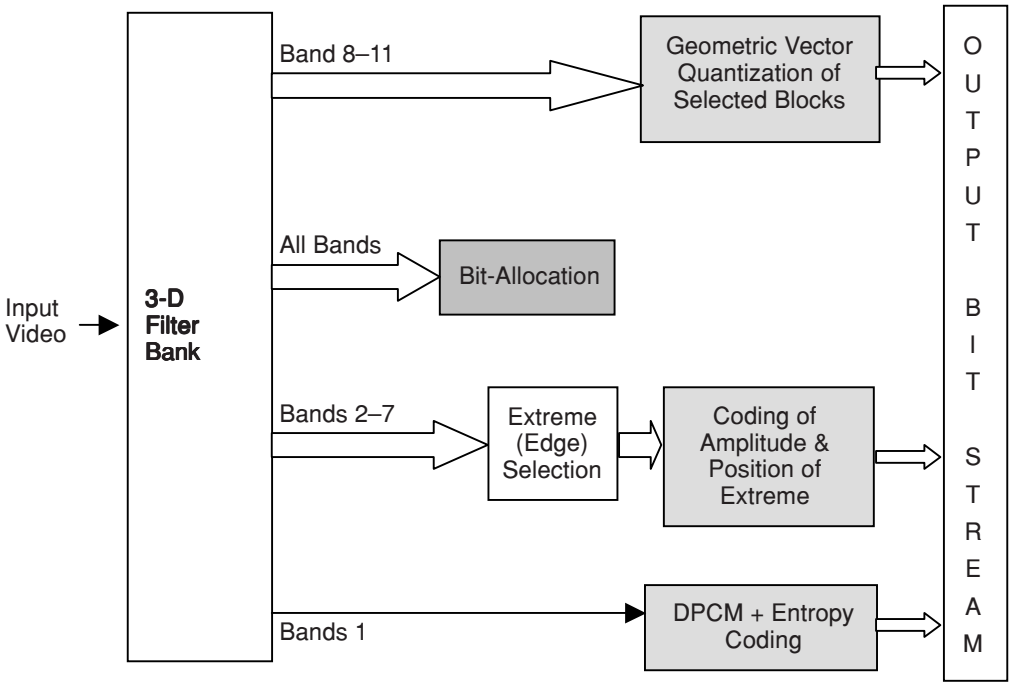


Figure 2 Block diagram of SMER Coder

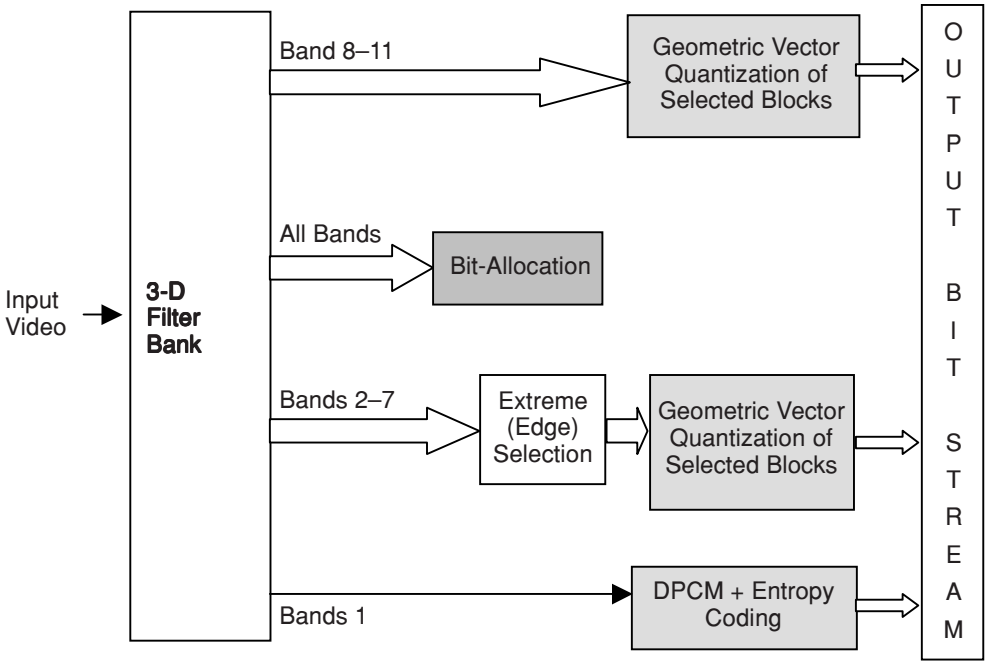


Figure 3 Block diagram of SMEG Coder

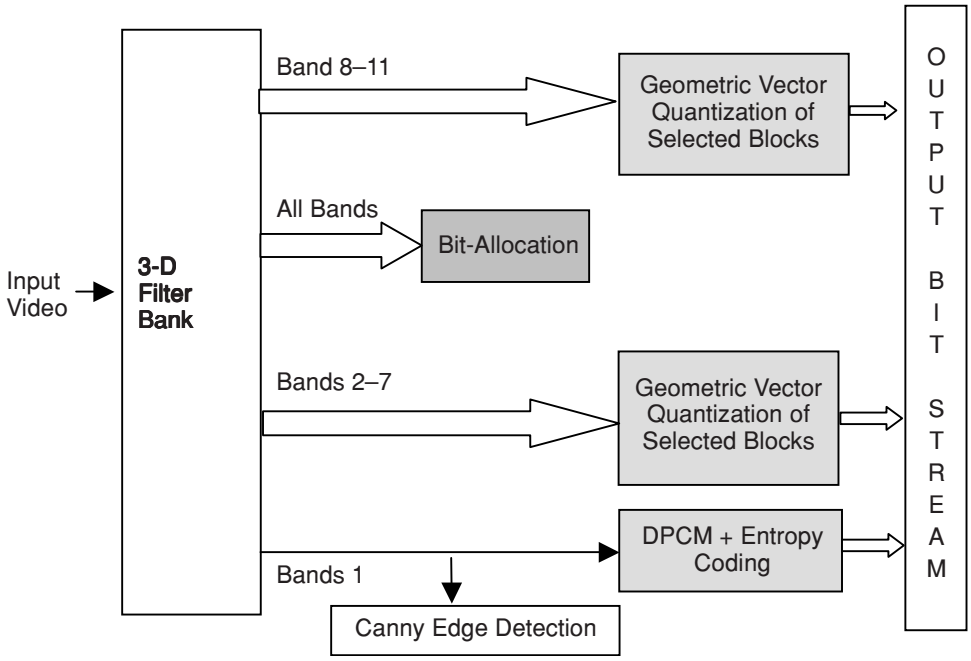


Figure 4 Block diagram of SCEG Coder

- **SMER**: System with Multiscale Edge detector and Reconstruction from extreme (Figure 2).
- **SMEG**: System with Multiscale Edge detector and Geometric vector quantization (Figure 3).
- **SCEG**: System with Canny Edge detector and Geometric vector quantization (Figure 4).

3.2.1 Geometric Vector Quantization

Each spatio-temporal frequency subband has a geometrical characteristic structure associated with it. For example, subband 2 (in Figure 1), corresponding to high-vertical/low-horizontal spatial frequency components consists of mostly horizontal edges or bars, whereas subband 3, corresponding to low-vertical/high-horizontal spatial frequency components, consists of mostly vertical edges or bars.

Geometric Vector Quantization (GVQ) is a vector quantization method where the codevectors are inspired by these edge related features of the high-frequency subbands [10, 14]. Unlike traditional vector quantization, GVQ does not depend on a training set. The codevectors for two-level GVQ are composed of binary-valued blocks reflecting the basic shapes found in the upper subbands and two intensity values, which are also transmitted for each coded block. Figure 5 shows geometric codebook consisting of nine 3×3 binary codevectors chosen for representing subband

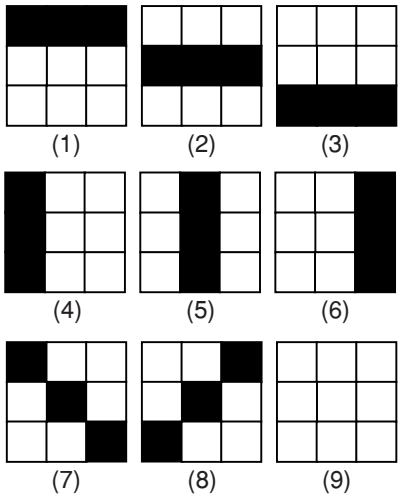


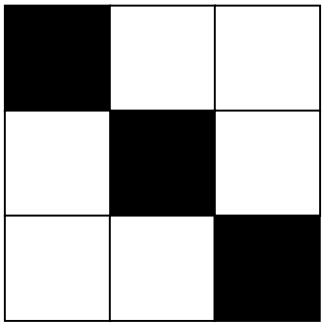
Figure 5 GVQ codebook of nine 3×3 code vectors

data. The codevector shapes are strips or strip combinations of various orientations and thickness. For a given input vector, an adaptive procedure finds the two intensities of each codevector to maximize the match to the input. This procedure is repeated for each codevector in the codebook, and the intensities with the best match are used to reproduce the input image block. The transmitted information includes the index of the chosen vector in the codebook and the two chosen output intensity levels for the areas indicated in black and white of the selected output (as shown in Figure 6), respectively, also an entropy coding is implemented on these information.

To find the intensity values for a vector of length $\boldsymbol{N}$, the input vector is divided into non-overlapping regions $\boldsymbol{B}$ and $\boldsymbol{W}$, corresponding to the black and white parts

x_1=1.2	x_2=14.2	x_3=1.2
x_4=18.7	x_5=2.6	x_6=19.3
x_7=23.0	x_8=12.9	x_9=3.9

(*a*) Input Vector



(*b*) Candidate Code vector

Figure 6 Example of implemented GVQ block matching

of the candidate codevector $\boldsymbol{k}$. An example of matching the input vector to one of the candidate vectors is illustrated in Figure 6. The two intensity values ($\boldsymbol{L_{1,k}}$ and $\boldsymbol{L_{2,k}}$) could be found based on averaging the gray level of correspondent pixels means, (N_B and N_w are number of black and white pixels)

$$L_{1,k} = \frac{1}{N_B} \sum_{i=1}^{N_b} x_i \tag{4}$$

$$L_{2,k} = \frac{1}{N_W} \sum_{i=1}^{N_w} x_i \tag{5}$$

The candidate codevector with these intensity values then compared with the input image vector, and a distance D_k is determined as given by

$$D_k = \sum_{i=1}^{N_b} (x_i - L_{1,k})^2 + \sum_{i=1}^{N_w} (x_i - L_{2,k})^2 = D_{1,k} + D_{2,k} \tag{6}$$

For the example shown in Figure 6, the calculation is as follow

$$L_1 = \frac{1}{3}(x_1 + x_5 + x_9) = 2.5 \tag{7}$$

$$L_2 = \frac{1}{6}(x_2 + x_3 + x_4 + x_6 + x_7 + x_8) = 19.7 \tag{8}$$

$$D_7 = 3.64 + 90.79 \approx 94 \tag{9}$$

Similar calculation should be done for other candidate to find related distance ($\boldsymbol{D_k}$, $\boldsymbol{k = 1, 2, 3, \ldots M}$), where $\boldsymbol{M}$ is the size of the codebook ($\boldsymbol{M = 9}$ in Figure 5). These parameters shown in Table 3, which determines that the candidate vector shown in (Figure 6) is the best since $\boldsymbol{D_7}$ the lowest one.

Table 3 Approximate distance of codevectors from input sample vector

Approximate Distance								
D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	D_9
175	158	162	147	182	139	94	183	153

4.0 RATE CONTROL

In 3-D subband coding the frames are all encoded in similar way and there is no I, P and B frames like hybrid coding scheme. Each time two consecutive frames are

analyzed and 11 subbands are created using them. Based on this fact, it is not possible to drop some frames in order to control output bit rate and the only tools for controlling output bit-rate are quantization schemes. The system can work in two modes of operation, open-loop and closed-loop. In open-loop mode the system works based on setting the number of quantization level for DPCM coder of lowest band and scalar quantizer for extreme amplitude (in the first system) and amplitude of intensities in GVQ (for second and third system). In closed mode of operation, based on operating bit, the bit-allocation module set these parameters based on allocated bit-rate for each one (assuming they follow a logarithmic distribution), later based on requested total output bit-rate, for every 2 subband decomposition (3 consecutive frames which means almost 400 ms) , the output bits are buffered and in case of overflow some high frequency bands could be eliminated.

5.0 COMPLEXITY OF SYSTEM

In this section a rough approximation of number of multiplication are provided as a major of system complexity. The complexity of a 3-D subband coder could be split to coding and filtering part. The size of spatial filters (Table 1) are all 4, and they have symmetric or anti-symmetric characteristic. Table 4 shows this calculation for each of analysis filter and input frame with QCIF format. Encoding high frequency subbands with GVQ is a process of calculation of Euclidian distance (based on Equation 13). Based on the size of codebook this means a total of $9 \times (9 + 2) = 99$ multiplication for each vector quantization. The third column in Table 4 shows this

Table 4 Approximate number of multiplications in developed coders

Subband	Multiplications	
	Filtering	**Encoding**
11	1.5×144×176×2	99×{(72×88)/9}=69696
10	144×88×2	69696
9	1.5×144×176×2	69696
8	144×88×2	69696
7	1.5×144×176×2	69696
6	144×88×2	69696
5	1.5×144×176×2	69696
4	144×88×2+1.5×72×88×2	17424
3	72×44×2	17424
2	1.5×72×88×2	17424
1	72×44×2	–
Total	(analsysis:456192)+(synthesis: 456192)+ (encoding: 540144) =1452528	

calculation for encoding subbands. As it shows totally the amount of multiplications (two filtering + encoding) for the two systems of SMEG and SCEG are roughly 1.5 Million floating operations. This amount of calculation for the first system is much higher due to its recursive analysis-synthesis process in its decoding process (that usually needs to be repeated for 2 or 3 times). This increase the number of filtering process to $4 \times 456192 \cong 1.8$ or $6 \times 456192 \cong 2.8$ million multiplications. (The other calculations are GVQ for higher subbands 8-11 and some few multiplications that could be roughly around 0.5 million multiplications).

6.0 PRACTICAL RESULTS AND COMPARISON OF SYSTEMS

The developed video coders were tested under different output bit-rates for all the three system and H. 263. Four bit-rates (128, 64, 32, 16 Kbps) were selected and results reported for five video sequences (Miss-America, Claire, Suzie, Salesman, Carphone), each one for the first 150 frames. Table 5 to 8 show average PSNR for different coders and H.263 standard. Also variation of PSNR for two sequences of Miss America and Suzie has been illustrated in Figure 7 and 8.

Table 5 Average PSNR for in Proposed Systems and H263 for 16 Kbps

	PSNR (dB)			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	33.4	33.3	31.2	34.0
Ms. America	35.3	35.2	33.0	35.9
Salesman	28.2	28.0	24.5	28.8
Suzie	26.3	26.1	24.4	26.9
Carphone	26.3	26.0	24.7	27.1

Table 6 Average PSNR for in Proposed Systems and H263 for 32 Kbps

	PSNR (dB)			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	35.5	35.3	33.3	36.0
Ms. America	37.5	37.4	35.1	38.1
Salesman	30.3	30.1	27.9	31.1
Suzie	31.9	31.4	27.9	32.5
Carphone	30.2	30.0	27.8	31.0

Table 7 Average PSNR in Proposed Systems and H263 for 64 Kbps

	PSNR (dB)			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	39.4	39.2	38.0	39.8
Ms. America	41.5	41.2	39.8	41.9
Salesman	32.4	32.0	30.7	33.0
Suzie	35.6	35.0	33.7	36.1
Carphone	32.3	31.8	30.0	32.9

Table 8 Average PSNR in Proposed Systems and H263 for 128 Kbps

	PSNR (dB)			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	41.0	40.6	39.5	41.4
Ms. America	43.3	42.9	41.8	43.6
Salesman	35.7	35.3	34.2	36.3
Suzie	37.1	36.8	35.3	37.6
Carphone	34.8	34.3	33.1	35.3

In order to further compare the quality of systems, at first a pair comparison subjective test , based on ITU-T P.910, with three levels, (Better, The same, Worse), for comparing the SMER and H.263 at 16 Kbps were organized [15]. Table 9 shows these results. The average score is 0.15. In fact in most cases (65%), viewer preferred to choose "the same" option. This could be justified as similar performance of systems, however to have a more clear investigation some other objective evaluation methods were used. Based on this, three parameters of ANSI T1.801.3 defined

Table 9 Subjective test on comparison of two coder (H.263, SMER)

	Score (range [–1 , 1])						
Bit-rate Kbps	Claire	Miss. America	Sales	Suzie	Carphone	**MOS**	**STD**
16	0.15	0.25	0.15	–0.15	0.20	0.1200	0.1615
32	0.10	0.25	–0.20	0.25	0.15	0.1100	0.1851
64	-0.15	0.10	0.15	–0.10	0.15	0.0300	0.1440
128	0.10	0.15	–0.10	0.15	–0.05	0.0500	0.1173

measures that are mostly recommended for evaluation at low bit-rate were used [16] (Refer to appendix A. for more a brief information on this standard). These parameters are "maximum added motion energy" (T_1), and "maximum lost motion energy" (T_2), and "average edge energy difference" (S_3). Using the definition, mentioned in Tables 10 and 11, these parameters at all four output bit-rates are calculated and provided in Tables 12 to 23. It should be noted that for ANSI parameters, an absolute smaller value means less distortion or better performance for systems.

6.1 Discussion on Results

As it is expected at higher bit-rates (128 Kbps) the difference between different systems is negligible. At low bit-rate, with using entropy coder for lowest frequency subband, the results shows much better performance compared to previous reported experiments [10–11]. In fact both system of SMER and SMEG have competitive result compared to H.263 system. Meanwhile clearly the performance of both systems is better than SCEG.

In terms of PSNR, the results show that, both two systems (SMER, SMEG) completely following H.263.The difference between the proposed coders (SMER and SMEG) and H.263 has been always less than 0.7 dB. This difference shows itself more in Salesman and Carphone video sequence with an average of 0.7 dB.

Table 10 Spatial Impairment Parameters Defined in ANSI T1.801.3

No.	Parameter Name	$a_{in}(t)$	$a_{out}(t)$	Definition
S_1	Maximum added edge energy	SI_{stdev}	SI_{stdev}	$\min_t\left(\frac{a_{in}(t)-a_{out}(t)}{a_{in}(t)}\right)$
S_2	Maximum lost edge energy	SI_{stdev}	SI_{stdev}	$\max_t\left(\frac{a_{in}(t)-a_{out}(t)}{a_{in}(t)}\right)$
S_3	Average edge energy difference	SI_{stdev}	SI_{stdev}	$rms_t\left(\frac{a_{in}(t)-a_{out}(t)}{a_{in}(t)}\right)$
S_4	Maximum HV-to-non HV edge energy difference	$\left(\frac{HV(t)+0.5}{\overline{HV(t)}+0.5}\right)$	$\left(\frac{HV(t)+0.5}{\overline{HV(t)}+0.5}\right)$	$\max_t\left(\frac{a_{in}(t)-a_{out}(t)}{a_{in}(t)}\right)$

Table 11 Temporal impairment Parameters Defined in ANSI T1.801.3

No.	Name Parameter	$a_{in}(t)$	$a_{out}(t)$	Definition
T_1	Maximum added motion energy	TI_{ms}	TI_{ms}	$\max_t \left\{ \log \frac{a_{out}(t)}{a_{in}(t)} \right\}$
T_2	Maximum lost motion energy	TI_{ms}	TI_{ms}	$\min_t \left\{ \log \frac{a_{out}(t)}{a_{in}(t)} \right\}$
T_3	Average motion energy difference	TI_{ms}	TI_{ms}	$rms_t \left(\log \frac{a_{out}(t)}{a_{in}(t)} \right)$
T_4	Maximum lost motion energy with noise removed	$TI^2_{rms} -$ min (TI^2_{rms})	$TI^2_{rms} -$ min (TI^2_{rms})	$rms_t \left(\frac{a_{in}(t) - a_{out}(t)}{a_{in}(t)} \right) pp$
T_5	Percent repeated frames	–	TI_{ms}	$\left(\frac{\#\ frames(a_{out}(t) < Threshold}{total\ frames} \right).100\%$
pp: positive part, index t : operation is over whole frame				

Table 12 Maximum added motion energy (T_1) of Coders at 16 Kbps

	T_1			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	2.0325	2.0984	2.3745	2.0497
Ms. America	1.3917	1.3992	1.7862	1.4790
Salesman	1.3630	1.3973	1.4633	1.4618
Suzie	1.3918	1.4554	1.5848	1.6774
Carphone	1.4447	1.4534	1.5745	1.4568

Table 13 Maximum added motion energy (T_1) of Coders at 32 Kbps

	T_1			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	1.8477	1.9120	2.0577	1.9898
Ms. America	1.1772	1.1819	1.4015	1.2326
Salesman	1.2407	1.3406	1.5711	1.3553
Suzie	1.2519	1.2640	1.3673	1.3015
Carphone	1.2636	1.2697	1.3647	1.3063

Table 14 Maximum added motion energy (T_1) of Coders at 64 Kbps

	T_1			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	1.8092	1.8132	1.8637	1.8345
Ms. America	1.1683	1.1695	1.2893	1.1752
Salesman	1.0395	1.0772	1.2461	1.1065
Suzie	1.2451	1.2837	1.3429	1.2976
Carphone	1.2285	1.2832	1.5055	1.2854

Table 15 Maximum added motion energy (T_1) of Coders at 128 Kbps

	T_1			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	1.7965	1.8479	1.9549	1.8480
Ms. America	1.1552	1.1654	1.3185	1.1692
Salesman	0.9891	0.9983	1.0682	1.0095
Suzie	1.2312	1.2413	1.3964	1.2977
Carphone	1.2237	1.2546	1.3112	1.2651

Table 16 Maximum lost motion energy (T_2) at 16 Kbps

	T_2			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	0.9058	0.9059	1.0459	0.9109
Ms. America	0.4223	0.4243	0.6629	0.4370
Salesman	0.6933	0.6953	0.8232	0.70386
Suzie	0.5934	0.6094	0.6658	0.6082
Carphone	0.0104	0.1062	0.2008	0.1063

Table 17 Maximum lost motion energy (T_2) at 32 Kbps

	T_2			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	0.6057	0.6082	0.8584	0.6353
Ms. America	0.1815	0.2098	0.4183	0.2254
Salesman	0.6123	0.6291	0.7157	0.6576
Suzie	0.4977	0.5095	0.7192	0.5618
Carphone	0.1581	0.1595	0.1760	0.1774

Table 18 Maximum lost motion energy (T_2) at 64 Kbps

	T_2			
Video Seq.	SMER	SMEG	SCEG	H.263
Claire	0.4923	0.5040	0.6166	0.5256
Ms. America	0.1396	0.1561	0.1733	0.1705
Salesman	0.5420	0.5602	0.5898	0.5753
Suzie	0.4687	0.4890	0.6498	0.5033
Carphone	0.1941	0.2132	0.4375	0.2123

Table 19 Maximum lost motion energy (T_2) at 128 Kbps

	T_2			
Video Seq.	**SMER**	**SMEG**	**SCEG**	**H.263**
Claire	0.4500	0.4909	0.4503	0.4944
Ms. America	0.1179	0.1582	0.3143	0.2019
Salesman	0.4874	0.4986	0.5699	0.5004
Suzie	0.4495	0.4940	0.5190	0.5052
Carphone	0.1958	0.2303	0.3294	0.2514

Table 20 Average edge energy difference (S_3) at 16 Kbps

	S_3			
Video Seq.	**SMER**	**SMEG**	**SCEG**	**H.263**
Claire	0.081	0.0880	0.0881	0.0724
Ms. America	0.064	0.0720	0.0923	0.0577
Salesman	0.101	0.1022	0.1037	0.0984
Suzie	0.169	0.1740	0.1853	0.1652
Carphone	0.066	0.0735	0.0931	0.0626

Table 21 Average edge energy difference (S_3) at 32 Kbps

	S_3			
Video Seq.	**SMER**	**SMEG**	**SCEG**	**H.263**
Claire	0.020	0.0285	0.0496	0.0169
Ms. America	0.035	0.0403	0.0616	0.0255
Salesman	0.060	0.0681	0.0721	0.0546
Suzie	0.166	0.1699	0.1848	0.1590
Carphone	0.038	0.0476	0.0496	0.0292

Table 22 Average edge energy difference (S_3) at 64 Kbps

	S_3			
Video Seq.	**SMER**	**SMEG**	**SCEG**	**H.263**
Claire	0.008	0.0083	0.0132	0.0011
Ms. America	0.034	0.0435	0.0591	0.0272
Salesman	0.037	0.0441	0.0661	0.0308
Suzie	0.164	0.1705	0.1680	0.0308
Carphone	0.034	0.0385	0.0415	0.0340

Table 23 Average edge energy difference (S_3) at 128 Kbps

	S_3			
Video Seq.	**SMER**	**SMEG**	**SCEG**	**H.263**
Claire	0.0078	0.0111	0.0196	0.0060
Ms. America	0.033	0.0392	0.0355	0.0252
Salesman	0.030	0.0339	0.0322	0.0231
Suzie	0.163	0.1672	0.1763	0.1629
Carphone	0.033	0.0393	0.0529	0.0291

The three ANSI parameters, provides an understanding of spatial and temporal distortion in reconstructed video sequence. In terms of temporal parameters (T_1 and T_2) the proposed systems show 10 to 20 percent lower values (that means better performance), but in terms of spatial parameter (S3), the developed system show 12 to 30 percent higher values (that means lower performance), this means that developed coder is creating more blurriness in spatial domain but on the other hand the bit-allocation module forces 3-D subband coder to lose its spatial coding performance to keep encoding high temporal subbands during fast motions, and make the encoder performance to be more stable. This fact also shows itself in Figure 7 and 8 by comparing the variation of PSNR in different systems. This variation for 3-D

Figure 7 Block diagram of the III Proposed Systems

Figure 8

subband system has been lower compared to H.263 system (exact computation in this two cases, shows that the standard deviation of PSNR in SMER and SMEG is around 10%–15% lower than H.263).

In terms of complexity, the second proposed system (SMEG) is clearly outperforms the H.263 system, as it has only around 2 million multiplication and also it

does not have any feedback structure and high storage because of small codebooks. As a measure of comparison, it should be mentioned that benchmark hybrid coders, need around 20–100 times more than this amount of operation, mainly because of motion estimation process [9].

7.0 SUMMARY

A video-coding scheme based on three dimensional subband coding is described. The three proposed schemes have competitive results with traditional low bit-rate coding system like H263. Since the bit-stream contain information of edge of the video sequence the proposed system is suitable for application like tele-monitoring and where fast detection of motion of objects is crucial.

ACKNOWLEDGEMENTS

This work was supported by the University Technology Malaysia research grant 72112.

REFERENCES

[1] Kovacevic, K., M. Vetterli. 1994. Wavelet and subband coding. Englewood Cliffs, NJ: Prentice Hall.
[2] Karlsson, K., M. Vetterli. 1987. Subband coding of video signals for packet switched networks. *SPIE Int. Conf. on Visual Communication and Image Processing*, 446–456.
[3] Ashourian M. 2001. *Low Bit-rate 3-Dimensional Subband Video Coding*, Ph.D. Thesis, University Technology Malaysia.
[4] Mallat S., S. Zhong. 1992. Characterization of signals from multiscale edges. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 14(7):710–732.
[5] Mallat, S. 1990. Zero-crossings of a wavelet transform. *IEEE Transaction on Information Theory*. 37 (6):1019–1033.
[6] Chang, S.G., M. Vetterli. 1997. Spatial adaptive wavelet thresholding for image de-noising. *IEEE Int. Conf. on Image Processing (ICIP97)*. 374–377.
[7] Misiti M. *et. al.* (1999). Matlab, Wavelet Toolbox, User's guide. USA-Massachusetts: The MathWorks.
[8] Jayant, N. S. 1988. Digital coding of waveform, principle and application to speech and video. 3th. Ed. Englewood Cliffs, N.J.: Prentice-Hall.
[9] Lundmark, H., R. Forchheimer. 1994. Image sequence coding at very low bit-rates: A review. *IEEE Transaction on Image Processing*, 3(5):589–616.
[10] Podilchuk, C., N. Jayant, N. Farvardin. 1995. Three-dimensional subband coding of video. *IEEE Transaction on Image Processing*. 4(2):125–139.
[11] Tham, J.Y.; S. Ranganath, A. Kassim. 1998. A highly scalable wavelet-based video codec for very low bit-rate environment. *IEEE Journal on Selected Areas in Communications*. 16(1)12–27
[12] Ngan, K.N. *et. al.* 1994 . Very low bit rate video coding using 3-D subband approach. *IEEE Transactions on Circuits and Systems for Video Technology*. 4(3):309–316
[13] Chooi, W.L. *et. al.* 1994. 3-D subband coder for very low bit rates. *IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP94)*. 405–408.
[14] Cosman, P. C., R. M. Gray, M. Vetterli. 1996. Vector quantization of image subbands: A survey. *IEEE Transaction on Image Processing*, 5(2):202–225.
[15] International Telecommunication Union. 1999. Subjective video quality assessment methods for multi-

media applications. Geneva: (ITU-T P.910).
[16] American National Standard Institute. 1996. Digital transport of one-way video signals - parameters for objective performance assessment. Washington: (ANSI T1.801.03).
[17] Wolf, S. 1997. Measuring the end-to-end performance of digital video systems. *IEEE Transaction on Broadcasting*. 43(3):320–328.
[18] Wolf, S., M. H. Pinson, S. D. Voran. 1991. Objective quality assessment of digitally transmitted video. *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*. 477–482.
[19] Yu, A., R. Lee, M. Flynn. 1997. An evaluation of video fidelity metrics. *IEEE Int. Conf. on Computer and Communication (Compcon97)*. 49–55.

APPENDIX A. ANSI PARAMETERS

ANSI T1.801.3 is a standard that classify major impairments in video signals and propose methods for measuring them [16]. Several types of parameters known as scalar, vector and matrix parameters are defined. In following sections these parameters are explained briefly. In general scalar parameters have more successful in quality measurement of video signals at low bit-rates Scalar, or one dimensional, features produce one value per video sequence. ANSI has defined a set of scalar quality parameters that can be extracted from the source and destination video. Two useful topics that are used in the computation of several scalar quality parameters are “spatial information” (SI), and “temporal information” (TI), which are defined at first.

Spatial Information (SI) can be derived by using different edge-enhancing filters. Sobel filters are suggested in the ANSI paper [16]. The edge enhanced images are obtained by linear convolving each video frame luminance ($\boldsymbol{Y_n}$) with the matrix of Equation (A.1) and Equation (A.2).

$$H_v = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad [A.1]$$

$$\boldsymbol{H_h} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad [A.2]$$

The results from these convolutions are called horizontal and vertical spatial information of the $\boldsymbol{n^{th}}$ video frame,

$$SI_v(n) = Convolution \ \ (Y_n, H_v) \quad [A.3]$$

$$SI_h(n) = Convolution \ \ (Y_n, H_h) \quad [A.4]$$

Magnitude (radius) and phase of the spatial information are defined as:

$$\boldsymbol{SI_r}(\boldsymbol{n}) = \sqrt{\boldsymbol{SI_h^2} + \boldsymbol{SI_v^2}} \qquad [A.5]$$

$$\boldsymbol{SI_\theta}(\boldsymbol{n}) = \arctan\left(\frac{\boldsymbol{SI_v}(\boldsymbol{n})}{\boldsymbol{SI_h}(\boldsymbol{n})}\right) \qquad [A.6]$$

The standard deviation of $\boldsymbol{S_r}$, named $\boldsymbol{SI_{stdev}}$, is the major statistical character for spatial information which used in calculation of ANSI spatial parameters. Table 10 shows the list parameters defined based on spatial information. For each parameter the feature related to source frame (at the time t) is mentioned by $\mathbf{a_{in}(t)}$ and the one related to output video (compressed one) by $\mathbf{a_{out}(t)}$. The four parameter S_1, S_2, and S_3 are directly calculated based on $\boldsymbol{SI_{stdev}}$. The term $\boldsymbol{HV}$ in S_4, is the average of $\boldsymbol{S_r}$ value over all pixels which belongs on an edge and are oriented within a predefined angle of vertical and horizontal; The $\overline{\boldsymbol{HV}}$ is its complement, means $\boldsymbol{S_r}$ values for pixels which are edges in other angles. For each parameter a physical interpretation is mentioned in Table 10. Parameters based on spatial information can be interpreted as indicator for added or lost edges in the destination scene compared to the source scene. Added edges result from impairments such as tiling, error blocks, and noise. Lost edges can result from impairments like blurring [16–19].

Temporal Information(TI): Describes the difference (movements) between two adjacent frames, $\boldsymbol{Y_{n-1}}$ and $\boldsymbol{Y_n}$ as defined below

$$TI(n) = Y_n - Y_{n-1} \qquad [A.7]$$

As Table 11 shows, the RMS (root mean square) value of $\boldsymbol{TI}(\boldsymbol{n})$ is main feature for calculating the 5 scalar parameters (T_1 to T_5). Quality parameters based on temporal information can be interpreted as indicating added or lost motion in the destination scene compared to the source scene. Added motion result from impairments such as jerkiness, error blocks, and noise. Frame repetition obviously introduces lost motion.

There are some facts that should be considered in using ANSI parameters, first is that as their definition shows, they are not independent from each other. The second thing is that depends on application and coding bit-rate, some of these parameters might be more important than others in evaluation and comparison of system. The experiments in [16–18] which are a combination of subjective tests and evaluation of ANSI parameters show that specially at low bit-rate, or lower resolution and quality, our visual system shows sensitivity more to average distortion in spatial domain (S_3) and worth distortion in temporal domain (T_1,T_2).