Jurnal Teknologi, 34(D) Jun 2001: 63–82

PYRAMID VECTOR QUANTIZATION OF VIDEO SUBBAND WITH DITHERING

MOHSEN ASHOURIAN[1], ZULKALNAIN MOHD. YUSOF, SHEIKH HUSSAIN S. SALLEH, SYED ABD. RAHMAN S. A. BAKAR

Abstract. This paper describes the development of a low complexity and fixed-rate video compression scheme based on three-dimensional subband coding of video signals. The video codec first performs three-dimensional subband decomposition on a group of video frames, and then encode high frequency subbands with pyramid vector quantization and lowest tempo-spatial band with a DPCM coding in time and space. To improve the visual quality of reconstructed video, different types of subtractive and non-subtractive dithering of pyramid vector quantizers were experimented and its effectiveness was proved by a standard pair comparison subjective test. Coder complexity was reduced by using longer filters in the first level of spatial decomposition for better selectivity and coding gain and shorter filter in the second level of decomposition for lower complexity. Results at different low bit-rate (64, 128 and 384 Kbps) for several standard video sequences are reported and compared with ITU standard H.263.

1.0 INTRODUCTION

Based on application and design, there are different types of video compression systems. In this paper, we explain our results on developing a video compression system with low and constant bitrate using subband coding techniques that could be used without buffering and channel coding methods for fixed rate channels with low bit-error rate.

There are two kinds of redundancy that exist in a video sequence, namely temporal and spatial redundancies. In many existing video coding systems, spatial redundancies are removed by using techniques such as subband coding or DCT (discrete cosine transform) coding techniques, and temporal redundancies are reduces by predictive coding in conjunction with motion estimation-compensation techniques [1]. However, the recently introduced method of three-dimensional subband coding has shown successful redundancy reduction for both spatial and temporal redundancy reduction with much lower complexity [2–5]. The major challenge in subband coding design is selection of proper filter banks for decorrelating information in subbands and optimum selection of quantizers for different subbands based on their statistical characteristics. In the proposed video coder, pyramid vector quantization

[1] Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Skudai, Johor Darul Ta'zim, Malaysia. E-mail: mohsena@ieee.org

(PVQ) used for compression of high frequency subbands, and DPCM for lowest frequency band, and proper modifications for improving their performance are provided.

The rest of this paper is organized as follows. Section 2 discusses the first stage of our video coder, which is a three dimensional filter bank. In section 3, characteristics of video subband are explained. The complete block diagram of the propesed system and different encoding methods used for different subbands are explained in section 4. Section 5 explains results of system implementation and finally section 6 summarizes the works and concludes the paper.

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 [3–5].

Figure 1 shows the specific 3-D subband framework chosen, which consists of 11 spatio-temporal frequency bands. The terms HP and LP refer to high-pass and low-pass filtering, where the subscripts t, h, and v refer to temporal, horizontal, and vertical filtering respectively. The temporal frequency decomposition is restricted to only two subbands using Harr filters [1, 3]. This means high-pass and low-pass temporal frequency bands are produced by the difference and average between 2 consecutive frames. Also using, longer filters and more channel decomposition might exploit better long-term correlation among consecutive video frames, which could result in better coding gain. In practice almost all reported 3-D subband video coder [3–5], it has been preferred to limit the number of channel decomposition to two, and their filters to short Harr filters due to these facts:

(1) Low complexity in implementation. To figure out his matter, consider a 3-D filter bank with 4 channel of temporal decomposition, even with one level of decomposition in spatial domain, this process will increase the number of subbands to be coded from 11 in Figure 1, to 20. This matter will increase the complexity by the addition of filtering stage.

(2) Shortages of available bit-rate. At low bit-rate, most of the available bit-rate should be allocated to the lowest frequency band (band 1 in Figure 1) to keep a minimum quality. Therefore, in practice, most of the times the coder is forced to drop high frequency (temporal or spatial) bands (or can say quantize them with zero bit). Based on this, there is no need of increasing the number of subbands.

(3) Potential delay problems. Using longer temporal filter means that, for filtering of any frame, the three previous frame of video sequence should be kept in memory which means a delay of four frames time.

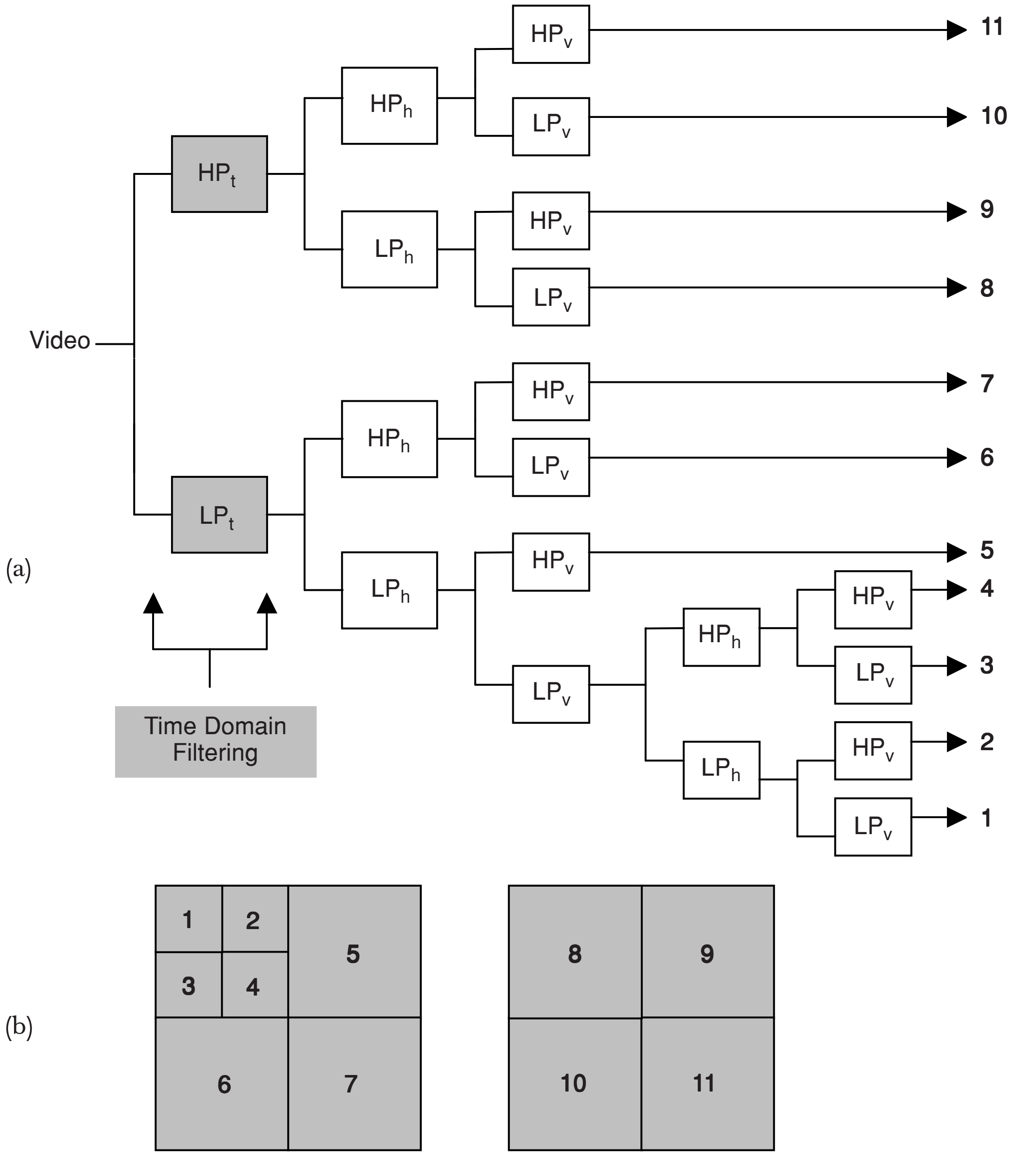

Figure 1 Selected 3-D Filtering ((*a*) Structure, (*b*) Frequency Map)

(4) Recucing error robustness performance. Using longer temporal filters will result in using more number of previous frames to take part in the construction of a new frame. This will increase long-term dependency of encoded bit-stream, which can result in more susceptibility to channel noise and probability of error propagation.

(5) Undesired visual artifacts. Similar to ringing effect caused by long filter in spatial filtering, in temporal domain, energy of consecutive frames compared to

each other could change because of ri pples in the step response of temporal filters.

(6) Coding gain. Only lowest tempo-spatial subband has high correlation in time and this redundancy has been exploited in developed system with a DPCM coding of it. Therefore, using longer temporal filters will not result in further improvement of coding gain.

In case of spatial filters, the lowest temporal band has decomposed two times (bands 1 to 7), but high temporal band only once (bands 8 to 11). In fact, since in low bit-rate application, video sequences do not have high spatial details or fast motion (or even if they have, ignoring them is not important) more number of temporal or spatial subbands are not necessary and it would be useless, since there is not enough available bit for allocating to them.

A lot of investigation has been done on the selection of spatial filters [6]. Based on coding efficiency, longer filters are usually preferred because they are sharper in frequency domain, but ringing effect around image edges and higher complexity in hardware implementation counterfeit the advantages [2]. In this paper, the proposed video coder uses a PVQ for encoding subbands where Johnson's 12 coefficient filters were selected [7], since the coding gain in PVQ highly depends on decorrelation of information by filter bank, that is, depends on selectivity of filters [8–10]. The first 6 coefficients of Johnson's low pass analysis filter are shown in Table 1. Since it is a symmetric filter and belongs to QMF filter bank family, the other three filters of analysis and synthesis part can be derived from it [7].

Table 1 Low pass analysis filter (Johnston 12) in fixed rate • system

N	LPF Analysis
1	–0.003809699
2	0.018856590
3	–0.002710326
4	–0.084695940
5	0.08846992
6	0.484389400

3.0 CHARACTERISTIC OF VIDEO SUBBANDS

The eleven subbands of Figure 1 could be classified based on their temporal and spatial frequency decomposition as follows:

1. Low temporal and spatial freguency band (Band 1)
2. Low temporal and high spatial frequency bands (Bands 2 to 7)
3. High temporal and low spatial frequency band (Band 8)
4. High temporal and high spatial frequency bands (Bands 9 to 11)

Figures 2 shows samples of these subbands for Salesman sequence. Table 2 shows normalized average energy of frames for whole Salesman dan Claire sequence and

(*a*) Band 1

(*b*) Band 2

(*c*) Band 8

(*d*) Band 9

Figure 2 Samples of Subbands

Table 2 Average and standard deviation of subbands energy for Salesman

Band No.	Average	Standard deviation
1	8.8875	0.2621
2	0.3139	0.181
3	0.3851	0.0193
4	0.0815	0.0049
5	0.2426	0.0136
6	0.2458	0.0096
7	0.0240	0.0015
8	12.3916	0.2053
9	0.2426	0.0136
10	0.2458	0.0096
11	0.0240	0.0015

(*a*) Band 1

(*b*) Band 2

(*c*) Band 8

(*d*) Band 9

Figure 3 Samples of histograms of subbands

standard deviation of this energy (in comparison of energy, it should be noted that the size of subbands 1 to 4 is 1/4 of others). Figure 3 shows the histogram of some subbands amplitude. Based on these results, some major facts about characteristics of these subbands are summarized as follows:

(1) Band 1 blurred version of the original frame and has much higher energy compared to others and automatically like image coding the most visual importance [5].
(2) Bands 2 to 7 are information of texture and sharpness of signal in spatial frequency domain. The energy of these bands depands on the amount of the information scene. Among these bands clearly the bands 5 and 7 has much lower energy since they are the results arts of two time highpass filtering (vertical and horizontal).
(3) Band 8 has higher average energy compared to other high temporal bands. However since it has direct relation to movement of object in scene-during the time, its variance is also high. In a 3-D scheme, that there is no motion estimation modelu. It's energy could be used as a measure of amount of movement in frames in bit-allocation module [3].
(4) Band 9–11 have low energy, but high variation in time. They represent sharp and fast movements of objects in scene. However in low bit-rat coding these matters are very rare in scene and usually ignored because of shortage of available bit-rate.
(5) Amplitude histograms for this sequence and other reported investigations [3–5] shows that like image subbands, band 1 does not follow any distribution but bands 2–11 follow well a generalized Gaussian distribution.

Selection of optimum quantizer for different subbands based on their statistical characteristics and visual importance is the key factor for developing subband coder. In the following sections the block diagram of system and different quantizers used for different bands are explained.

4.0 BLOCK DIAGRAM OF PROPOSED SYSTEM

Figure 4 shows the block diagram of the designed fixed-rate video coder. At first, the signal is passed through a 3-dimensional filter bank. For the lowest frequency subband DPCM coding is used and for the high frequency subbands PVQ were used. Based on the percentage of compression, or output bit-rate of system, bit-allocation module set the parameters of quantizers. In order to improve visual quality at low bit-rates, a dither signal is added to high frequency subbands. In the following sections, different modules of system and its results will be explained in more details.

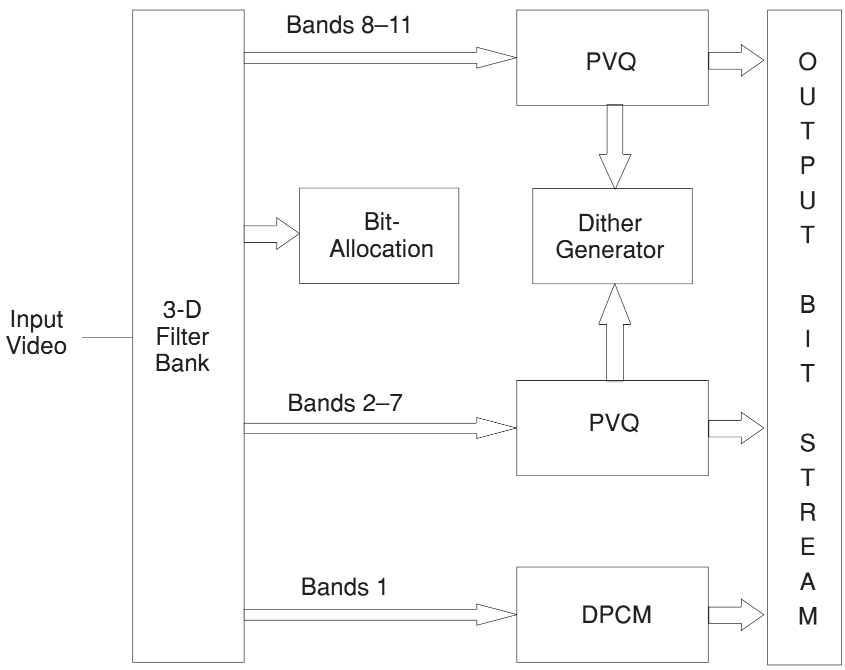

Figure 4 Block diagram of fixed-rate coding system

4.1 Lowest Tempo-Spatial Subband

High energy and visual importance makes the lowest tempo-spatial frequency subband (Band 1) very important in image and video coding. Since the distribution of this subband is highly image dependent and does not follow a fixed statistical distribution, quantization scheme which are based on assumption of a fixed statistical distribution for source, such as Lloyd-Max scalar quantizer or pyramid vector quantizer could not be used directly. Since error occurring in this subband tends to have stronger impact on the overall reconstructed image quality than those occurring in the higher frequency subbands, sclar quantization is usually preferred to vector quantization. In the event of bit error, no error propagation will occur, and only few pixels might be affected [3].

In contrary to high frequency subbands, the correlation properties of the lowest frequency band, both in time and spatial domain are high, which makes DPCM an efficient scheme for this band.

The process of DPCM is simple. Instead of coding the original signal $x(i, j, t)$ at position (i, j) and at times t, its difference from a predicted value $(x_p, (i, j, t))$ is coded;

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

The following linear predictive coding (LPC) strategies were tested on subbands data for several different image sequences

$$x_p(i,j,t) = e_1 x(i,j-1,t) + e_2 x(i-1,j,t) + e_3 x(i-1,j-1,t) \tag{2}$$

$$\begin{aligned} x_p(i,j,t) &= e_1 x(i,j-1,t) + e_2 x(i-1,j,t) + e_3 x(i-1,j-1,t) \\ &\quad + e_4 x(i,j,t-1) + e_5 x(i,j-1,t-1) + e_6 x(i-1,j,t-1) \\ &\quad + e_7 x(i-1,j-1,t-1) \end{aligned} \tag{3}$$

$$x_p(i,j,t) = e_1 x(i,j-1,t) + e_2 x(i-1,j,t) + e_4 x(i,j,t-1) \tag{4}$$

$$x_p(i,j,t) = e_1 x(i,j-1,t) + e_2 x(i-1,j,t) \tag{5}$$

$$x_p(i,j,t) = e_4 x(i,j,t-1) \tag{6}$$

where $e_1, e_2, \ldots e_7$ are prediction coefficients and are calculated using Shur-Levinsor algorithm [3]. In contrary to tradition LPC coding of speech, the variation of prediction coefficients in different sequence and in one sequence from frame to frame is quite low (around 10%) [4], therefore by averaging, the fixed set of coefficient as follows, were selected.

$$\boldsymbol{e_1 = e_2 = e_5 = e_6 = 1/2;\ e_4 = 1;\ e_3 = e_7 = 1/4;} \tag{7}$$

Setting fixed coefficient has also the advantage that it is not necessary to transmit coefficients. The prediction gain, for were calculated for each frame based on this formula [1],

$$G_p = 10 \log_{10} \left\{ \frac{\sigma_x^2}{\sigma_d^2} \right\} \tag{8}$$

and then averaged over all frames in sequence. Here σ_x^2 is the variance of signal and the σ_d^2 is the variance $d(i, j, t$ in Equation [1]). The average of G_p over all frames are tabulated in Table 3 and 4.

The results in Table 3 shows, the second prediction scheme has the highest average coding gain for lowest frequency subband. This result is reasonable since more terms take part in this estimator (Equation 3) compared to other ones. Therefore a DPCM coding based on this method was used to improve its coding efficiency.

Table 3 Average prediction gain factor for lowest tempo-spatial band

	Gain (dB)				
Video Sequence	Equ. 5.2	Equ. 5.3	Equ. 5.4	Equ. 5.5	Equ. 5.6
Miss America	13	23	19	10	17
Suzie	10	21	12	9	11

Table 4 Average prediction gain factor for high spatial bands

	Gain (dB)				
Video Sequence	Equ. 5.2	Equ. 5.3	Equ. 5.4	Equ. 5.5	Equ. 5.6
Miss America	2	5	2	5	6
Suzie	3	1	2	4	5

4.2 Pyramid Vector Quantization

Different types of vector quantization have been tried for efficient coding of high frequency subbands in image and video coding [11]. Based on the distribution of high frequency subbands, which could be approximated well with generalized Gaussian distribution functions [8–9, 11]. Fischer introduced pyramid vector quantization for encoding these signals and proved that at high bit-rate its performance in coding is close to source entropy [8]. Another advantage of pyramid vector quantization is its fixed output bit-rate. Most of other vector quantizers only have good performance in encoding subbands if an entropy coding is added on to their output index, which makes them variable rate coder. In order to use these coders at fixed rate, it is necessary to devise a buffering scheme, which has difficulties in rate control and buffer overflow [8–12].

Figure 5 The four steps in polar pyramid vector quantization

Vector quantizer is used in the proposed system. Polar PVQ has a low complexity and regular encoding method, the basic quantization steps of a polar PVQ are illustrated for a 2-D input signal in Figure 5. As it shows these steps are:

(1) Calculating the vector radius, r, defined as the absolute norm of the vector to be coded.
(2) Projecting the vecotr to be coded onto the pyramid surface of radius $\boldsymbol{K}$ by scaling each vector element by $\boldsymbol{K/r}$. (The parameter $\boldsymbol{K}$ determines the number of lattice point on each shell, and has direct relation to selected output bit-rate).
(3) Quantizing the scaled vector to the nearest lattice point on the pyramid surface.
(4) Enumeration process, which means indentifying the index of the nearest lattice points.

Lattice index and the vector radius are parameters that should be transmitted. The proposed PVQ coder uses different vector dimension (or block size) for each subband based on the operating bit-rate of the system. The lattice radius $\boldsymbol{r}$ was quantized with a non-uniform scalar quantizer and the lattice indexes were enumerated with a magnitude enumeration method [8]. The following section explains about bit-allocation and selected bit-rates for quantization of lattice radius and lattice index in each case. In the first step of this design process the block size (or vector dimension in 1-D) for each band is determined. The second step is to find the maximum number of bits for lattice indices based on available bit-rate for subband. This determines the scaled lattice radius ($\boldsymbol{K}$ in Table 5). For higher percentage of coding bigger block size is better, but using a smaller block size, significantly improves the error resiliency by localizing the effect of possible bit errors [12]. Another major benefit of using small block size is significant reduction in hardware, a result of smaller indices and memory. With a larger vector dimension, a larger radius is required to maintain the same coding rate, which significantly increases the size of memory needed to store the tables. For example, for a given coding rate, to decode PVQ-encoded vectors of dimension 16 will require a memory size roughly eight times larger than that required for decoding vectors of dimension 4.

4.3 Bit-Allocation Among Subbands

The goal of bit-allocation is typically to minimize the overall distortion of the encoder subject to constraints such as a maximum overall bit rate. Also energy of a subband cannot be an exact measure of its visual importance, however because of computation simplicity, it is a common way to allocate the available rate based on it. Based on this fact, since bands 4, 7, 9, 10, and 11 have the minimum bit per pixel because of their low energy compared to their number of pixels, dropping them has less effect in visual quality of image. In fact subbands 4, 7 a contain information of highpass filtering in vertical and horizontal direction, and band 9, 10, 11 a contain information of high temporal and high spatial filtering. In a low bit rate application, where the high texture and very fast motions are not important, discarding these high frequency components are not visible [12]. However band 8 is kept and coded efficiently as it is the only band that shows the change in temporal domain. Only whenever its energy is lower than a threshold, (1/4 of the average energy of Band 8 in subband decompositions) it si assumed that there has not been any change in scene and its bit-rate is allocated to low temporal subbands.

The bit-allocation used in PVQ does not have high flexibility. The reason is that after setting the size of blocks in a subband, only integer and fixed number of choices exist for the number of lattice index (which determine the number of bit for it). Table 5 shows the number of lattice index for block size of 16, based on variation of size of lattice, As it is clear, the gap between the lattice indexes are mostly high, making the bit-allocation much less flexible. This restriction and the fact that generally the video scene in a low bit-rate application does not have so much change justifies the use of a fixed bit-allocation scheme.

In order to further reduce the output bit-rate, a method known as "Toggle Decimation" was used [13]. In this methods bands 2, 3, 5 and 6 are updated with half rate of updating bands 1 and 8. Figure 6 shows the operation of this method. For even frames, band 3 and 6 are transmitted and for odd frame band 2 and 5. The reason behind effectiveness of this method is geometrical similarities that exist between band 2 and 3 and band 5 and 6, and after image feature of our visual system, such that refreshment of one of them will compensate non-refreshment of other one, [13]. For quantizing energy and mean of each eleven subbands and the original frame eight bits were used. This means $(11 \times 8) + (11 \times 8) + (8 + 8) = 192$ bits per frame or

Figure 6 Toggling of subbands for even (a), and odd (b) frames

Table 5 Lattice index for block of 4×4

K: Scaled Lattice Radius	**N: Number of Lattice Point**	**Log_2 (N): Number of Bits for Indexing**
1	32	5
2	512	9
3	5472	13
4	44032	16
5	285088	19
6	1549824	21
7	7288544	23
8	30316544	25
9	113461024	27

$192 \times 7.5 = 1440$ bit per second as side information. The bit-allocation scheme are tabulated in Tables 6 to 8. The chosen bit-rate (= 62, 113 and 359 Kbps) were selected close to common telecommunication line standar bit-rates (64, 128 and 384 Kbps). For example frame rate could increase to 15 per second, toggle decimation for bands (2, 3 and 5, 6) could be ignored or other high frequency bands could be encoded (bands 4, 7, 9, 10, 11) with similar lattice indices.

Table 6 Quantization scheme at 62 Kbps

Band	**Block size**	**No. bits for lattice index**	**No. bits for lattice radius**	**Total Bit/s**
1	1×1	DPCM: 3		35640
2,3	4x4	5	7	8910
5,6	8x8	7	5	8910
8	8x8	7	5	8910
Side Inf.	$192 \times 7.5 = 1440$			
Total Bit-Rate	62370 + 1440 = 63810 bps = 62.31 Kbps			

Table 7 Quantization scheme at 113 Kbps

Band	**Block size**	**No. bits for lattice index**	**No. bits for lattice radius**	**Total Bit/s**
1	1×1	DPCM: 4		47520
2,3	4×4	23	7	22275
5,6	8×8	24	6	22275
8	8×8	24	6	22275
Side Inf.	$192 \times 7.5 = 1440$			
Total Bit-Rate	114345 + 1440 = 115785 = 113.07 Kbps			

Table 8 Quantization scheme at 355 Kbps

Band	**Block size**	**No. bits for lattice index**	**No. bits for lattice radius**	**Total Bit/s**
1	1×1	DPCM: 8		95040
2,3	2×2	23	7	89100
5,6	4×4	23	7	89100
8	4×4	23	7	89100
Side Inf.	$192 \times 7.5 = 1440$			
Total Bit-Rate	362340 + 1440 = 363780 = 355.25 Kbps			

4.4 Improving PVQ Performance With Dithering

At low bit-rate coding, quantization noise has high dependency on input signal, which results in high distortion [14–15]. The classical methods for overcoming this problem in PCM coding is dithering. It is an addition of signal called dither before quantizer and subtracting (or some not-subtracting) it after reconstruction in receiver. The first scheme called subtractive dithering and the second method non-subtractive dithering [14–15]. In this work, we used a random dither. In order to generate a random dither for PVQ, a random vector that has a uniform distribution in (–1/2, 1/2) are generated and then each elements are mapped to PVQ domain by $\boldsymbol{K/r}$ of related lattice block.

In general dithering, especially non-subtractive one, the peak signal to noise ratio (PSNR) does not improve clearly, and its results are more subjective. In order to justify this matter, a simple pair comparison subjective test based on ITU standard P.910 [16] were provided and the 20 viewer were asked to judge the quality of video frame among different combination of the three case of no-dithering, with subtractive dithering and with non-subtractive dithering. Video frames were shown by computer monitor and with size of (10 × 12 cm) and the viewer watched it from a distance of 60 cm (5 times the height of the frame) and asked to give a score based on Table 9. The experiments have been done on all video sequence (Claire, Miss. America, Suzie, Salesman, Carphone) for 3 bit-rate (≅ 62, 113, 362 Kbps). The results for each sequences plus mean opinion scores standard deviation of results were elaborated [17].

The reduction of 10 to 30 percent in this value, mostly in case of Suzie, Claire and Salesman, could be another justification for effectiveness of dithering.

Table 9 Rating scale used in PC test

No.	Expression
–1	Worse
0	The Same
1	Better

4.5 Complexity of System

The complexity of 3-D subband coding algorithm is directly related to quantization scheme and filtering process. The process of DPCM coding does not need any major calculation, since the prediction coefficient are integer and all power of two, so multiplication could be done with simple shift process. PVQ like other type of lattice vector quantizers has a low complexity [8]. In PVQ the main process is scaling blocks of data (which means almost one multiplication per pixel), and then enumeration that is only a lookup-table process. Based on these facts, the only component of system that could be considered for further reduction of complexity is filter bank. As Figure 1 shows, the total number of spatial filtering process are 18, (also the input to these filters in the second layer for the lowest temporal band has half size of original input). As explained in bit-allocation scheme, because of low energy, the subbands 4, 7, 9, 10 and 11 are always discarded.

Meanwhile it has been proved in scalar quantization of subbands that coding gain of a multistage filter bank is less sensitive to selectivity (sharpness in frequency domain) of its filters in the second level of decomposition compared to the first level [17]. This means that it is possible to use shorter filters for the second layer of low temporal subbands for generating the bands (the six filter before bands 1 to 4 in Figure 1). In order to select an optimum filter to replace the Johnston's 12 coefficient filter for this level, several filters with shorter length were examined. The lengths of test filters are all even (similar to Johnstons' 12), in order to have modularity in implementation. Table 16 shows the result of average drop in PSNR (with the abbreviation of DPSNR) for three different symmetric filters for bit-rate of 62 Kbps. The first filter in an bi-orthogonal spline wavelet with one filter of length 8 and the other 4 (Biro 3.3 in Matlab wavelet toolbox [18]), the second one is a set of filter with length 6, proposed by [6] and is chosen based on maximum coding gain, and finally the third one is db3 from Daubechies family with a length of 6 [18]. The average length of all filters is 6 and their analysis and synthesis filters are symmetric or anti-symmetric, so they have similar computational complexity. As the results shows the second family of filters, from [6], shows better performance compared to two other filters, and the PSNR of system compared to original systems only drops around 0.25 to 0.40 dB which is negligible compared to reduction of complexity. In the final proposed system, the spatial filters in the first stage uses Johnston'12 and the second state uses filters based on Table 10. It should be mentioned that further reduction of length of filters could result in high reduction of performance and is not reasonable.

Table 10 Spatial Analysis/Synthesis Filter Coefficient Used in Second Stage

N	LPF Analysis	HPF Analysis	LPF Synthesis	HPF Synthesis
1	0.02349918	0.023499183	0.023499183	–0.023499183
2	0.16056522	0.160565220	–0.160565220	0.160565220
3	–0.8398316	–0.625391050	–0.625391050	0.8398316
4	–0.8398316	0.625391050	–0.625391050	–0.8398316
5	0.16056522	–0.160565220	–0.160565220	–0.16056522
6	0.02349918	–0.02349918	0.02349918	0.02349918

Table 11 Number of multiplications in analysis filter

Subband	No. of Multiplications
8	144×176×6×6
6	144×176×6×6
5	144×176×6×6
4	144×176×6×6+36×44×3×3
3	36×44×3
2	36×44×3×3
1	36×44×3
9, 10, 11, 7	—
Total	3,687,552

Finally, total number of multiplications in analysis part (which is almost same as synthesis part) is calculated as a measure of complexity in Table 11. Some facts that are considered in these calculation is that all of the selected filters are symmetric or anti-symmetric, which reduce the number of multiplication for filtering to half of the size of filter, and some bands are sharing some parts of their filtering (bands 2 and 1, or 4 and 3), that is considered for the upper subband in Table 11.

5.0 EXPERIMENTAL RESULTS

We examined five different monochrome video sequence (Claire, Miss America, Suzie, Salesman, Carphone) having QCIF format (176 × 144 pixel) with 7.5 frame/s. The chosen bit-rate (52.2, 113 and 355 Kbps) were selected close to the common telecommunication line standard bit-rates (64, 128 and 284 Kbps). The PSNR for any single frame is calculated as

Table 12 Average PSNR in proposed system at and H.263 at 62 Kbps Rate

	PSNR (dB)	
Video Seq.	**Proposed System**	**H.263**
Claire	37.0	39.8 (62 kb/s)
Miss. America	38.5	41.9
Salesman	30.2	33.0
Suzie	33.5	36.1
Carphone	30.0	32.9

Table 13 Average PSNR in proposed system at and H.263 at 113 Kbps Rate

	PSNR (dB)	
Video Seq.	**Proposed System**	**H.263**
Claire	39.5	41.4
Miss. America	42.0	43.6
Salesman	32.8	36.3
Suzie	34.9	37.6
Carphone	32.2	35.3

Table 14 Average PSNR in proposed system at and H.263 at 355 Kbps Rate

	PSNR (dB)	
Video Seq.	**Proposed System**	**H.263**
Claire	42.1	43.0
Miss. America	43.8	45.2
Salesman	37.6	39.6
Suzie	38.6	40.3
Carphone	36.9	38.8

$$PSNR = 10 \log_{10}\left(\frac{255^2}{MSE}\right) \qquad [9]$$

where MSE is the mean square difference between the input and output frames. Tables 12 to 14 show the average PSNR for different bit-rates of some video sequence and compare it with H.263, ITU video coding standard [19].

5.1 Discussion on Results

The results in previous section show, the PSNR of proposed system compared to H.263 is around 1-3 dB lower. The distortion in 3-D coding is mainly in the form of spatial and temporal blurring resulted by sparse quantization at low bit-rate. Blurring shows itself more in Salesman and Carphone sequence, which have higher background details or faster motion. Figure 7(*a*) and 7(*b*) show a frame of original Salesman sequence and its compressed one, which show loss of some detailed spatial information. The blocking distortion, like hybrid coders, does not exist, and non-smooth changes and contouring because of sparse quantization have been eliminated by dithering process. This could be seen well in comparing Figure 7(*c*), 7(*d*) (in background of Claire frame), and Figure 7(*e*), and 7(*f*) (faces in Suzie frames). It

Figure 7 Samples of coded video at 62 Kbps.
(a) Original salesman, (b) Reconstruction of salesman, (c) Claire compressed without dithering, (d) Claire compressed with dithering, (e) Suzie compressed without dithering and (f) Suzie compressed with dithering.

should be mentioned that dithering does not show a significant improvement in PSNR. In case of subtractive dither, in some sequence such as Claire, 0.2 to 0.3 dB progress can be seen for subtractive dithering due to change in distribution of high frequency subbands, but in general the effect of dithering was not clear in this aspect.

5.2 Coder Performance in Noisy Channels

Table 15 shows the DPSNR (average drop in PSNR, means amount of drop in PSNR compared to original system in noise free environment) of the system over a memoryless binary symmetric channel (BSC) with bit error rates (BER) of 10^{-2}, 5×10^{-3}, 10^{-3}, 5×10^{-4}, 10^{-4}. It should be mentioned that in all of these experiments, the side information (Information about energy of subbands and signal) are assumed to be transmitted without error. The reason is high importance and low bit rate of these information ($\cong$ 1.4 Kbps), which makes it possible to use a robust channel coding method.

Table 15 Average DPSNR for video sequences in noisy channel at 112 Kbps

	DPSNR (dB)				
BER	Claire	Miss. America	Suzie	Salesman	Carphone
10^{-4}	1.3	1.2	1.0	0.9	0.9
5×10^{-4}	3.9	3.4	2.9	2.8	2.7
10^{-3}	6.2	5.9	4.2	4.4	4.2
5×10^{-3}	12.3	11.7	9.8	9.7	9.3
10^{-2}	15.1	14.7	13.8	13.7	13.0

The results in different BER shows the coder performance in noisy channel is decreasing from 1 down to 14 dB for different bit error rates. The comparison to standard H.263 system without any channel coding is competitive, since in H.263 with variable length coding, (even at very low BER *e.g.* 10^{-5}), the decoder/encoder could lose their synchronization because of one bit error which could leed to destruction of several consecutive frames (usually around 5–10 frame) until next forced synchronization is sent. Further improvement of system in noisy channel is possible by modifications of subband quantizers [20].

6.0 SUMMARY

We have described a video-coding scheme based on three-dimensional subband coding and pyramid vector quantization. In our study the scheme has revealed the appealing properties, such as high compression with good perceptual quality. Meanwhile it should be considered that with having a fixed rate output in our proposed systems, the system is able to work in channels with low bit error rate without channel coding overhead bit-rates.

ACKNOWLEDGEMENTS

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

REFERENCES

[1] Jayant, N. S. 1988. Digital coding of waveform, principle and application to speech and video. 3th. Ed. Englewood Cliffs, N. J.: Prentice-Hall.

[2] Kovacevic, K., M. Vetterli. 1994. Wavelet and subband coding. Englewood Cliffs, NJ: Prentice Hall.

[3] 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.

[4] Podilchuk, C., N. Jayant, N. Farvardin. 1995. Three-dimensional subband coding of ideo. *IEEE Transaction on Image Processing*. 4(2). 125 – 139.

[5] Eyvazkhani, M. 1997. Video coding using 3-D subband decomposition for channel error compensation on and ATM network. *IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP97)*. 3009 – 3012.

[6] Asae, S. O., T. A. Ramstad. 1991. Some fundamental experiments in subband coding of images. SPIE Int. Conf. on Visual Communication and Image Proceeding, 734 – 744.

[7] Johnson, J. D. 1980. A filter family designed for used in quadrature mirror filter banks. *IEEE Int. Conf. on Acoustic, Speech and Signal Processing (ICASSP80)*. 290 – 294.

[8] Fischer, T. R. 1986. A pyramid vector quantizer. *IEEE Transaction on Information Theory*. 32(7): 568–583.

[9] Yusof, Z. M., T. R. Fischer. 1996. An entropy-coded lattice vector quantizer for transform and subband image coding. *IEEE Transaction on Image Processing*. 5(2): 289 – 298.

[10] Yusof Z. M.; Fischer T.R. 1995. Subband coding using a fixed rate lattice vector quantizer. *IEEE Int. Conf. on Image Processing (ICIP95)*. pp. 101 – 104.

[11] 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.

[12] Hung, A.C.; Meng T.H.Y. 1994. Error resilient pyramid vector quantization for image compression. *IEEE Int. Conf. on Image Processing (ICIP95)*. 583 – 587.

[13] Matsuki, J., et.al. 1999. Moving picture encoding using wavelet transform and afterimage effect of human visual system. *IEEE Int. Conf. On Intelligent Signal Processing and Communication*, 609 – 613.

[14] Gray, R. M., T. G. Stockham. 1993. Dithered quantization. *IEEE Transaction on Information Theory*. 39(5): 805 – 812.

[15] Bennett W.R. (1948). Spectra of quantized signals. Bell System Technical Journal, 27(5): 446 – 472.

[16] International Telecommunication Union 1999. Subjective video quality assessment methods for multimedia applications. Geneva: (ITU-T P.910).

[17] Ashourin, M. .(2001. Low Bit-rate 3-Dimensional Subband Video Coding, Ph.D. Dissertation, University Technology Malaysia.

[18] Misiti, M. et. al. 1999. Matlab, Wavelet Toolbox, User's guide. USA-Massachusetts: The MathWorks.

[19] International Telecommunication Union. 1995. Video coding for low bit rate communication. Geneva: (ITU-T H.263).

[20] Ashourian, M. et. al. .(2001. Robust 3-D Subband Coder Based on Dithered Pyramid Vector Quantizer." submitted to IEEE Region Ten International Conference (TENCON2001), Singapore 2001.