We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Low Bit Rate Video Compression Algorithm Using 3-D Discrete Wavelet Decomposition

Awad Kh. Al-Asmari

Electrical Engineering Dept., King Saud University, Riyadh
Saudi Arabia

1. Introduction

Video applications, like video teleconferencing, video telephones, and advanced television (ATV), have given the field of compression and transmission of digital video signals a significant importance. It is expected that the advances in video compression technology will play a crucial role in the transmission and display of three-dimensional video signal.

A typical image, for example, of size 512x512 pixels with 8 bits per pixel (bpp) needs storage capacity of about 2 Mbits. A video sequence, on the other hand, with the same frame size with 30 frames per second and a channel transmission rate of 64 kilo bits per second (kbps) would take about 17 minutes of transmission time. The required transmission time would become unmanageable with the continuously increasing demand of image base application. You can't put enough of it over a telephone line and you can't squeeze it into the broadcast bandwidth of available channels[1]. Therefore, image and video compression algorithms became a necessity to store or transmit these images.

Data compression is the science of representing information in a compact form by exploiting the different kinds of statistical structures that may be present in the data[2]. This is to reduce the number of bits per sample while keeping the distortion constant[3]. There is a great deal of correlation between neighboring pixel values of an image. Therefore, removing such redundant information and transmitting only the new information (the changes) enables us to reconstruct the original image. For video signals, redundancy over time between successive images can also be eliminated.

There are two types of compression: lossless and lossy. In the lossless compression the original image can be retrieved without error, while for the lossy compression, the original image can't be retrieved without error; an image copy close to the original can be retrieved.

---

[1] Realtime video compression poses challenge to designers and vendors alike, Computer Design, vol. 32, no. 7, pp. 67-70 (Child, July 1993).

[2] Hybrid coding of images for progressive transmission over a digital cellular channel, CISST'99 International Conference on Imaging Science, Systems and Technology, Monte Carlo resort, Las Vegas, Nevada, USA, PIN 128C (Al-Asmari et al.,June 28 – July 1, 1999).

[3] Introduction to data compression, Morgan Kaufmann Publishers Inc., San Francisco, California (Sayood, 1996).

Motion pictures expert group (MPEG) video standard is the most prevalent and widely used for video compression[3-6]. Also, the MPEG is an application specific standard and different versions of MPEG (Such as MPEG-1, MPEG-2, MPEG-4, and MPEG-7) are available for different applications and bit rates. The basic algorithm for all these versions is the same and is very similar to the other video compression standards.

The proposed algorithm is based on temporal filtering of image sequences with short symmetric kernel filters (SSKFs)[7-8], which are well known for their simplicity. In this paper, we use four SSKFs filters each with 4-taps and with decimation factor of 4:1 instead of two SSKFs filters each of 2-taps and with decimation factor of 2:1 used in classical 3D – decomposition algorithms[7-8]. The temporal filtering removes the redundancy in temporal domain. On the other hand, the pyramid coding (PC) is used for subband decomposition in the spatial domain. The vector quantization (VQ) and the absolute moment block truncation code (AMBTC) will be used to encode the spatial domain subbands.

2. 3-D decomposition of the video sequence

Practical video compression relies on techniques to reduce the amount of data required to represent a video sequence without any appreciable loss of information that can affect the visual quality of image. Our aim in this paper is to introduce an algorithm that compress the video sequence with a reduced bit rate and highest fidelity while keeping the computational complexity to a minimum to allow for easy hardware implementation. The complexity of the proposed algorithm is less than the standard MPEG-1 and MPEG-2 algorithms.

Since our approach to the video compression problem is based on progressive transmission, the multiresolution representation of the signals is required. The idea of multiresolution is that a coarser approximation to the signal is refined step by step until the desired resolution is obtained. An elaborate discussion on 3-dimensional decomposition of the video signals, using pyramid method for spatial decomposition, is given in this section.

A group of video sequences (four frames each) is decomposed into subbands in the temporal and spatial domains. The temporal frequencies are restricted to four subbands by passing four consecutive frames through four band pass filters of 4-taps each and with decimation factor of 4:1. By applying pyramid decomposition on these temporal subbands, nine spatial subbands are produced. In the next sub-sections, the temporal and spatial domain decomposition will be discussed in more details.

2.1 Temporal frequency decomposition

The 4-tap Haar basis functions (SSKFs filters) are used for temporal frequency decomposition. These filters have no phase or amplitude distortion and belong to a class of perfect reconstruction filters. The coefficients of SSKFs filters used in this research are given

---

[4] Digital pictures representation, compression, and standard, Plenium Press (Netravali & Haskell, 1995).

[5] Image and video compression standards: algorithms and architectures, Kluwer Academic Publishers (Bhaskaran & Konstantinidides & Hewlett Packard Laboratories, 1996).

[6] Digital compression of still images and video, Academic Press (Clarke, 1995).

[7] Subband coding of video for packet networks, Optical Engineering, vol. 27, no. 7, pp. 574-586, (Karlsson & Vetterli, July 1988).

[8] Pyramid coding of video signals for progressive transmission, CISST'97 International Conference, pp- 392-398, (Dantwala et al., June 30 – July 3, 1997).

in Table 1. The frequency responses of these filters are shown in Figure 1. $H_0(e^{j\omega})$ is the low pass filter, $H_3(e^{j\omega})$ is the high pass filter, while $H_1(e^{j\omega})$ and $H_2(e^{j\omega})$ are the band pass filters. The 3-dB bandwidth for these filters is approximately $\pi/4$.

	Lowpass filter	Bandpass filters		Highpass filter
N	$h_0(n)$	$h_1(n)$	$h_2(n)$	$h_3(n)$
0	0.5	0.5	0.5	0.5
1	0.5	0.5	-0.5	-0.5
2	0.5	-0.5	-0.5	0.5
3	0.5	-0.5	0.5	-0.5

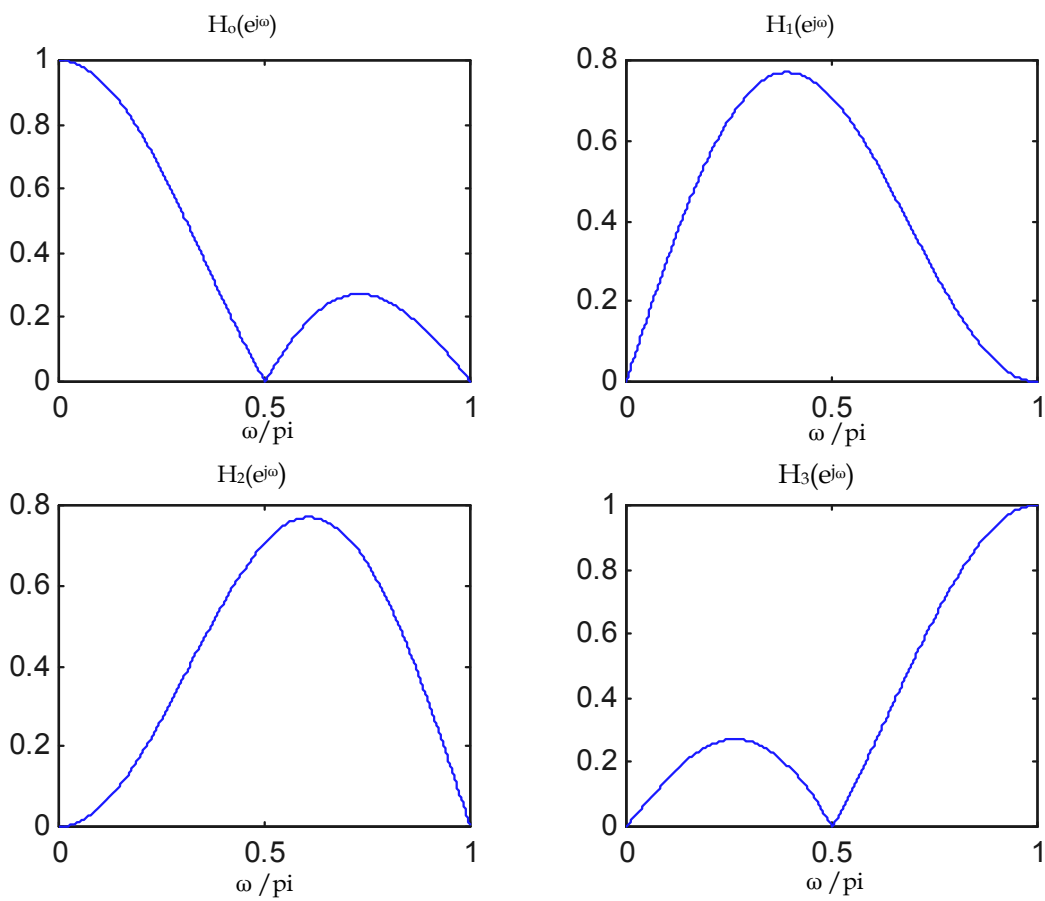Table 1. SSKFs Coefficients for Haar filters.

Fig. 1. Frequency response of 4-tap Haar filters.

The video sequence with four consecutive frames is decomposed in the temporal domain by passing these frames through the 4-tap Haar filters as shown in Figure 2.
The second four frames of Miss America sequence, that contains clear motion of the eyes, are filtered using SSKFs filters. These filters decompose the sequence into four temporal subbands. Namely, temporal low-low (T_{LL}), temporal low-high (T_{LH}), temporal high-low
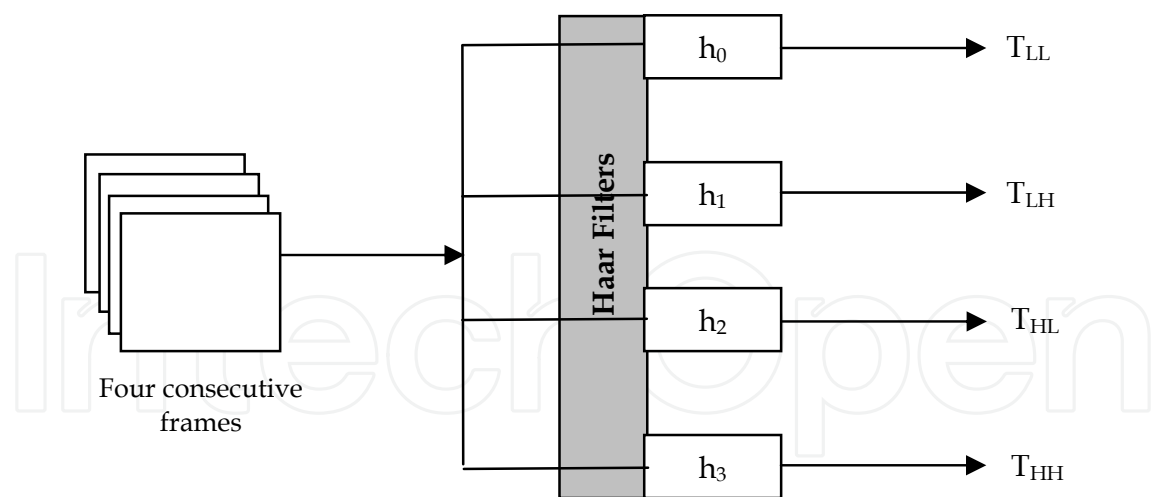
Fig. 2. Temporal filtering process using 4-tap Haar filters.

(T_{HL}) and temporal high-high (T_{HH}) subbands. Figure 3 shows the original and the filtered frames. It can be seen from this figure that the subband T_{LL} contains most of the image sequence information, while the subband T_{LH} contains most of the motion information in the four frames. The subbands T_{HL} and T_{HH} contain the edge information, which are relatively spares in the nature. Figure 3 (b) shows that the information of the temporal T_{LL} subband is highly correlated since it is only an average of the four frames, while the temporal (T_{LH}, T_{HL} and T_{HH}) subbands are of low correlated data. Compression is achieved since instead of transmitting four highly correlated frames of the video at 8 bpp each, only one frame with highly correlated data (T_{LL}) at high bit rate will be transmitted. The other frames (T_{LH}, T_{HL} and T_{HH}) with low correlated information will be transmitted with very low bit rate.

The average entropy is calculated for the original and the decomposed four frames. Table 2 shows the average entropy for the original and the temporal filtered frames of Miss America sequence. From this table, it is clear that the multiresolution decomposition process results in lower entropy than that of the original images. From Shannon's theory, the entropy for a signal can be minimized by decomposing this signal into sub-signals using orthogonal bases. Therefore, the calculated entropy is act as a lower measure for the information to be encoded[9].

Original frames	Entropy	Temporal frames	Entropy
Frame 4	5.66	T_{LL}	5.78
Frame 5	5.68	T_{LH}	2.39
Frame 6	5.66	T_{HL}	2.09
Frame 7	5.65	T_{HH}	2.11
Average	5.66	Average	3.1

Table 2. The average entropy for the four frames before and after the decomposition.

---

[9] A mathematical theory of communication, Bell System Tech. J.. Vol. 27, pp. 379-423, and pp. 623-656 (Shannon, July. 1948a, Oct. 1948b).
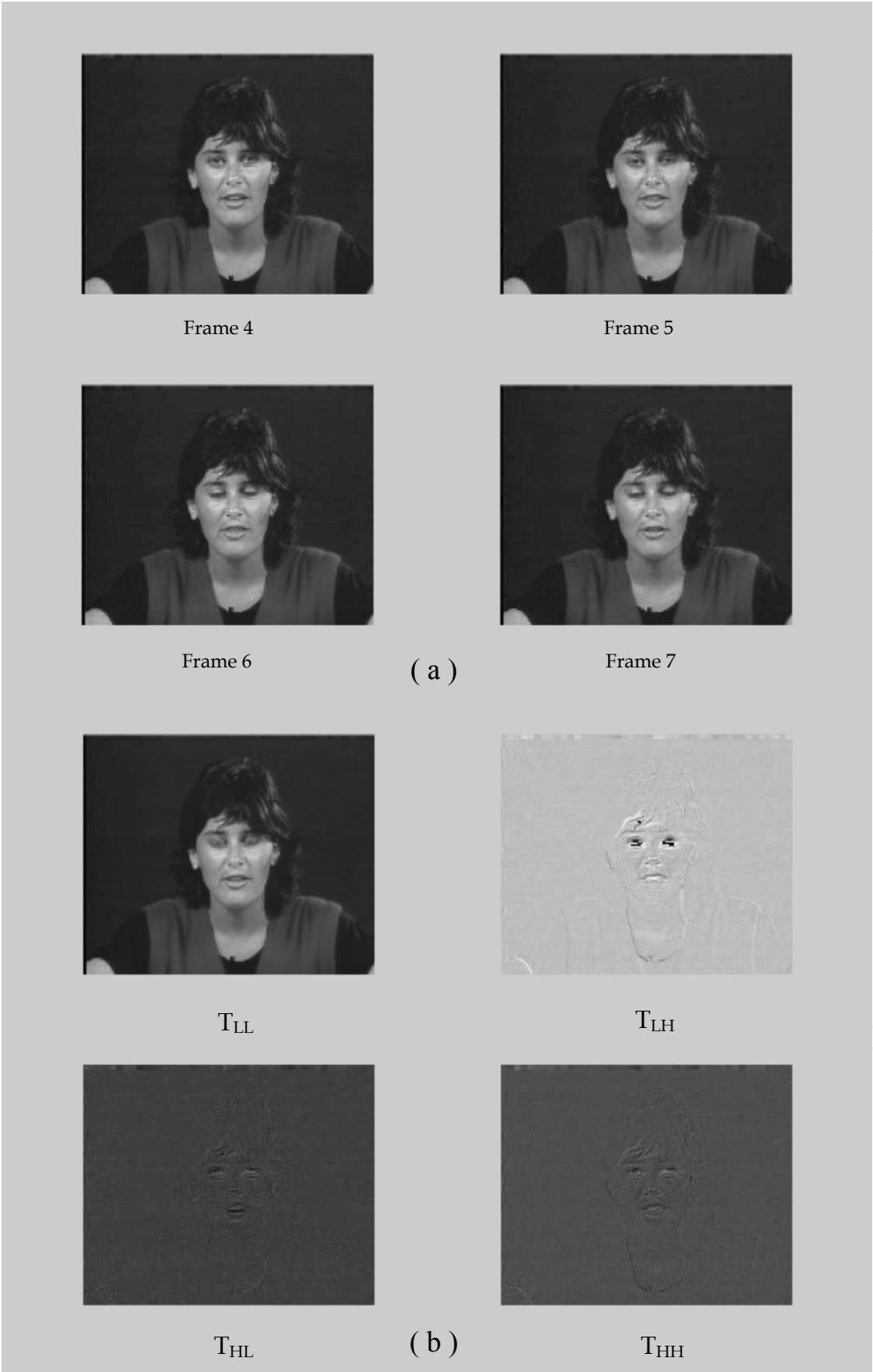
Fig. 3. (a) Original and (b) Temporal filtered frames of the second four frames of Miss America sequence.

2.2 Spatial domain decomposition

Spatial decomposition techniques aim at achieving data compression by discarding the redundant or visually non-perceptual information in an image. Subband coding (SBC) is one such scheme where the image spectrum is divided into a set of uncorrelated sub-spectra, and each of the sub-spectrums is treated individually depending on the amount of information it carries. The low-frequency bands with more energy content are given a higher priority as compared to the high frequency bands with low energy contents[10].

The four frames can be independently encoded using any image compression method available for still images. Since we are interested in multiresolution decomposition, pyramid decomposition method has been adopted for spatial domain decomposition. Pyramid coding (PC) has been found to be more robust to channel errors as compared to SBC[8]. This is so because neither data loss nor channel corruption in the error images has a serious effect on the quality of the reconstructed image. Moreover, lower bit rates are possible for PC as most of the energy of the pyramid structure lies in the lowest resolution baseband, which can be encoded at high bit rates to preserve maximum information without increasing the overall bit rate of the compressed image. Another advantage of the PC scheme is the ease of filter design as compared to SBC[10].

Subband filters are designed under the constraints of perfect reconstruction and narrow transition band. Both of these constraints are relaxed in PC as it makes use of bandpass images for reconstruction[10]. Because of these advantages, this research uses PC for spatial decomposition of images, instead of the SBC.

3. Pyramid coding

Burt and Adelson have suggested a method of pyramid coding that is suitable for progressive transmission[11]. In this method, the original image is filtered to be down-sampled by a factor of 2. The image thus obtained serves as a decimated image of the original. Then, the decimated image is filtered to be interpolated by a factor of 2 to have the same size of the original image. The difference between the original image and the interpolated one generates an error image. This is called the first level PC decomposition. This process can be further repeated over the decimated image to obtain higher levels. To achieve compression, the difference images and the decimated image are bit allocated depending on the amount of information in each subband. Those subbands with high information content will be assigned higher bit rate than those with lower information content.

3.1 24-tap filter

In [12], a 24-tap FIR filter has been introduced with a nominal bandwidth of $\pi/4$, which would allow a higher decimation factor than the conventional Gaussian filter used for pyramid coding[11]. The special property of this filter is that the passband ripples are designed for passband flatness. This ensures that the filter has as little in-band ripples as possible. It has a nominal bandwidth of $\pi/2M$ where M is the decimation factor (M = 4).

---

[10] Video Signal Transmission for IS-95 Environment, Electronic Letters, Vol. 36, no. 5, pp. 465-466 (Al-Asmari et al., 2nd March 2000).

[11] The Laplacian Pyramid as a compact Image Code, IEEE Trans Communication., vol. COM-31, no. 4 pp. 532-540, (Burt & Adelson, April 1983).

[12] Optimum Bit Rate Pyramid Coding with Low Computational and Memory Requirements, IEEE Trans. Circuit and Systems for video Tech., vol. 5, no. 3 pp. 182-192,(Alasmari, June 1995).

By using a pyramidal coding scheme which basically follows a rate of change of 4 instead of 2 as used in the conventional pyramid coding, the number of samples to be encoded are 20% less than the conventional pyramidal samples[12-13]. Another advantage of this filtering technique is that the 24-tap FIR filter involves 33% fewer computations as compared to the Gaussian filter when FFT algorithm is used. The third advantage of this filter is the lower entropy obtained when compared with that found when the Gaussian filter is applied[12].

4. The adopted spatial domain filtering techniques

The four temporal subbands (T_{LL}, T_{LH}, T_{HL} and T_{HH}) resulting from the temporal filtering process are further decomposed in the spatial domain using pyramid coding as shown in Figure 4. The temporal subband (T_{LL}) is further decomposed into three subbands using the
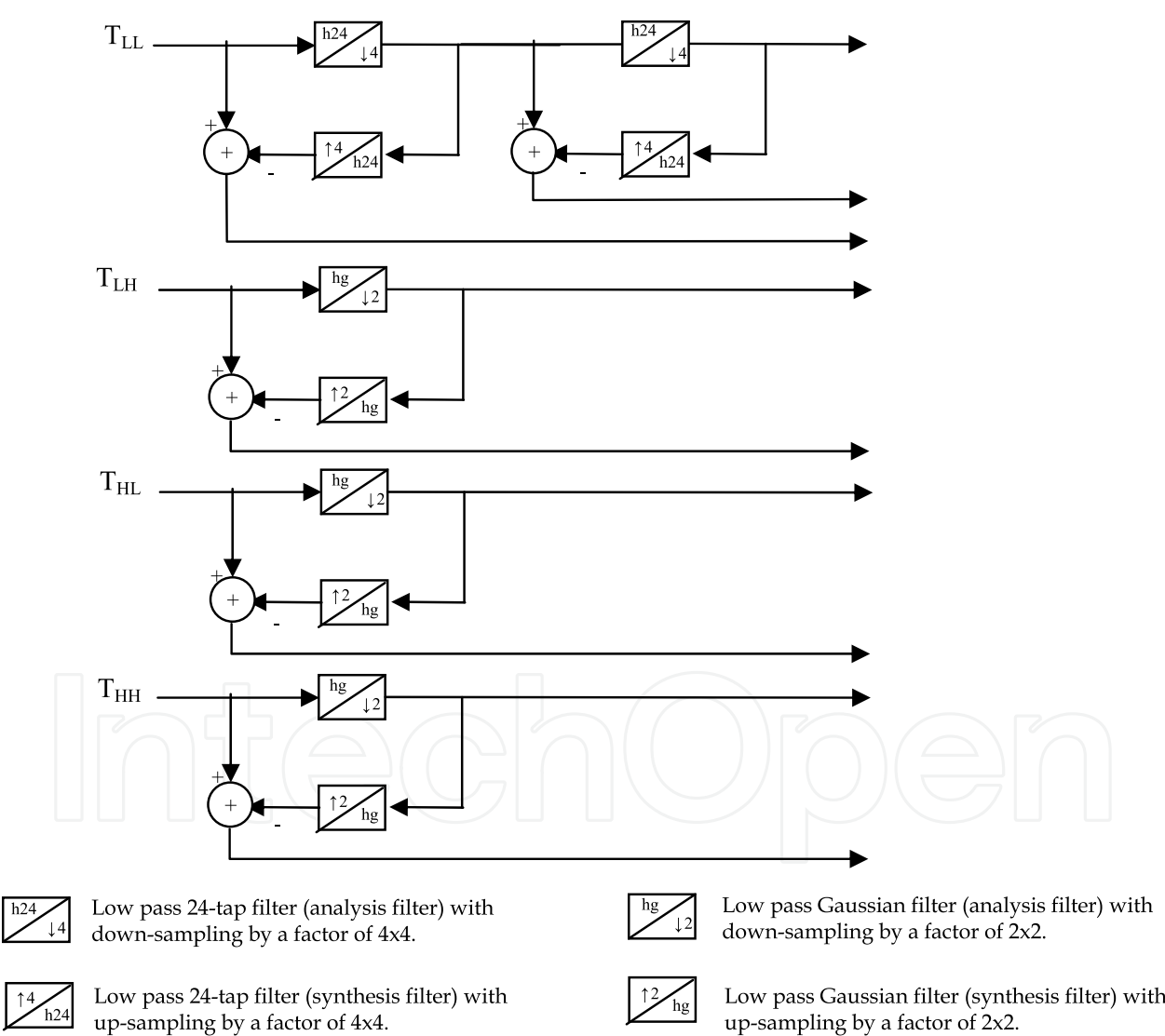

Fig. 4. Spatial decomposition for the temporal subbands using pyramid coding.

---

[13] Low complexity subband encoding for HDTV images, IEEE J. Select. Areas Commun., vol. 11, no. 1, pp. 77-87 (Coppisetti et al., Jan. 1993).

24-tap filter. It is found that the 24-tap filter will not give a good performance when used to decompose the temporal (T_{LH}, T_{HL} and T_{HH}) bands due to the spares nature of the information in these subbands[10]. The Gaussian filter given in[11] is used to decompose these subbands. Figures 5, 6, 7, and 8 show the spatial subbands for the temporal bands of Miss America sequence using the pyramid coding concept.

Fig. 5 shows the spatial pyramid decomposition of the temporal T_{LL} subband. Two levels of pyramid coding have been applied for this temporal subband. The decimated image (band 1), which contains most of T_{LL} subband information, is of dimension 18 × 22 pixels. The difference images of this band (band 2 and band 3) contain edge components and are of dimension 72 × 88 pixels and 288 × 352 pixels; respectively.
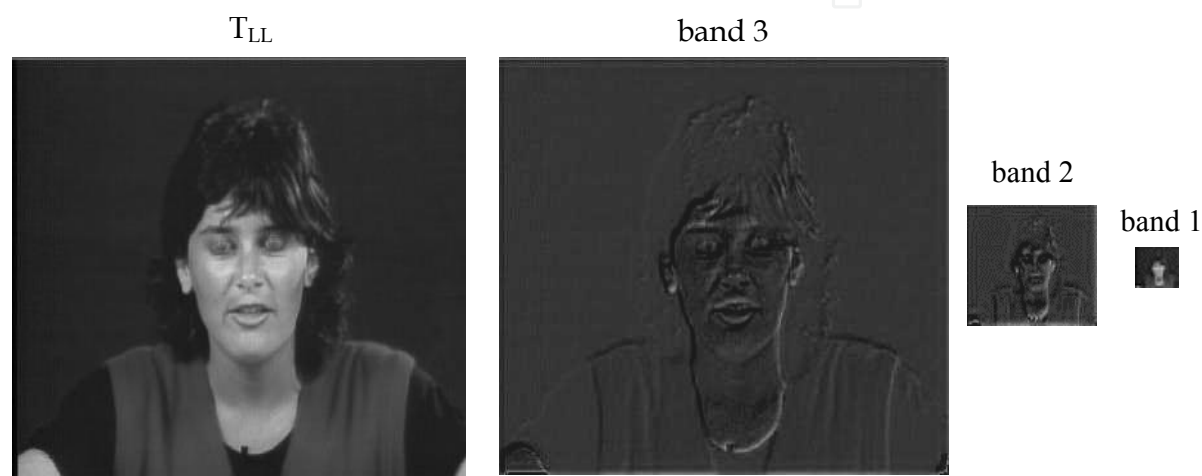

Fig. 5. Spatial subbands for the temporal low-low band.

The spatial pyramid decomposition of the temporal T_{LH} subband is shown in Figure 6. One level pyramid coding has been applied for this temporal subband. The decimated image of this band (band 4) is of dimension 144 × 176 pixels. The difference image (band 5) is of dimension 288 × 352 pixels.
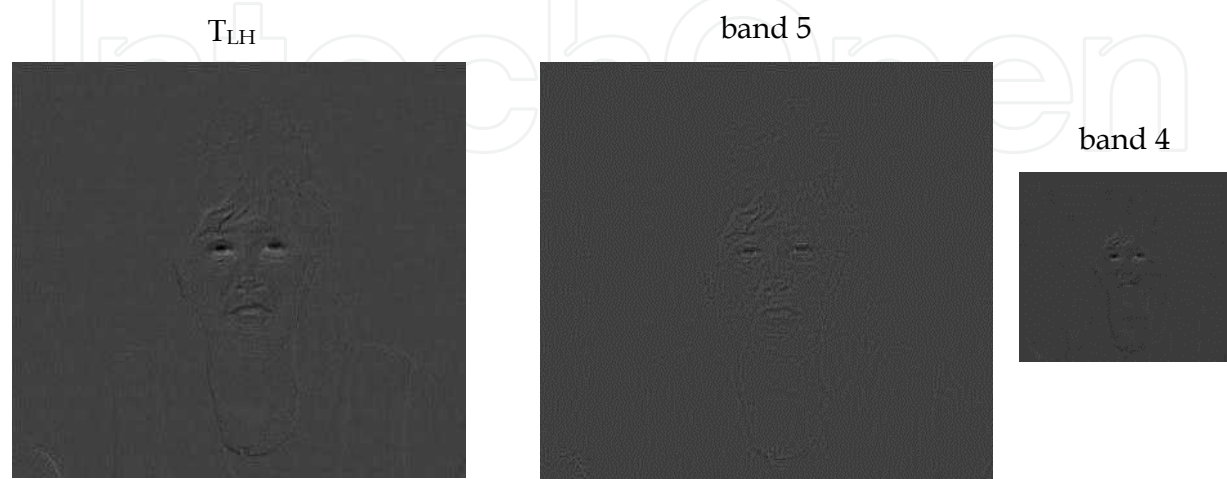

Fig. 6. Spatial subbands for the temporal low-high band.

In Figure 7 the spatial pyramid decomposition of the temporal T_{HL} subband is shown. One level pyramid coding has been applied for this temporal subband. The decimated image (band 6) is of dimension 144 × 176 pixels while the difference image (band 7) is of dimension 288 × 352 pixels.
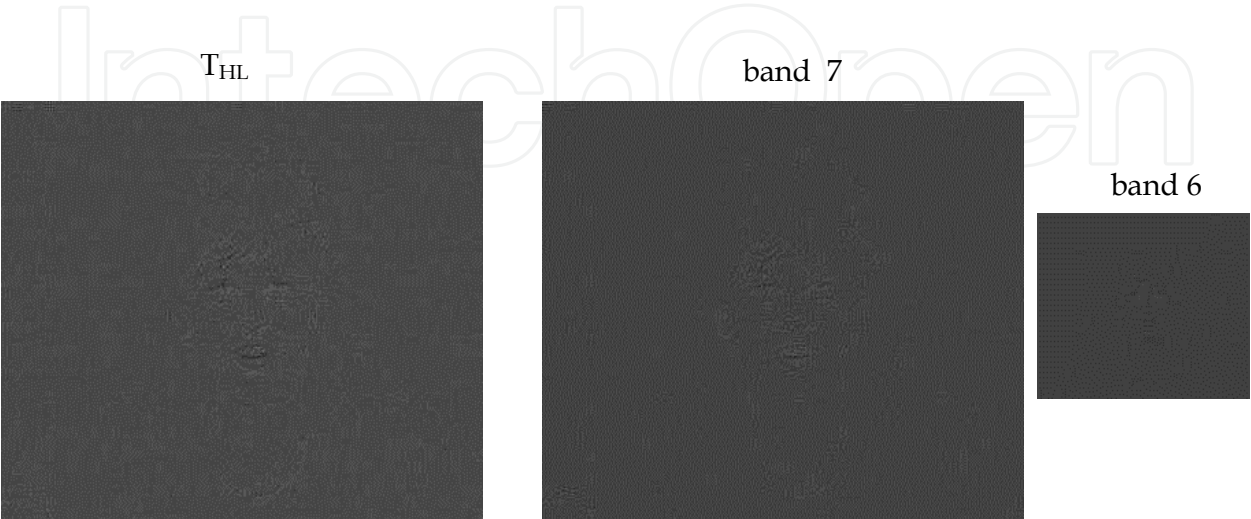

Fig. 7. Spatial subbands for the temporal high-low band.

Fig. 8 shows the spatial pyramid decomposition of the temporal T_{HH} subband. One level pyramid coding has been applied for this temporal subband. The decimated image (band 8) is of dimension 144 × 176 pixels and the difference image is of dimension 288 × 352 pixels.
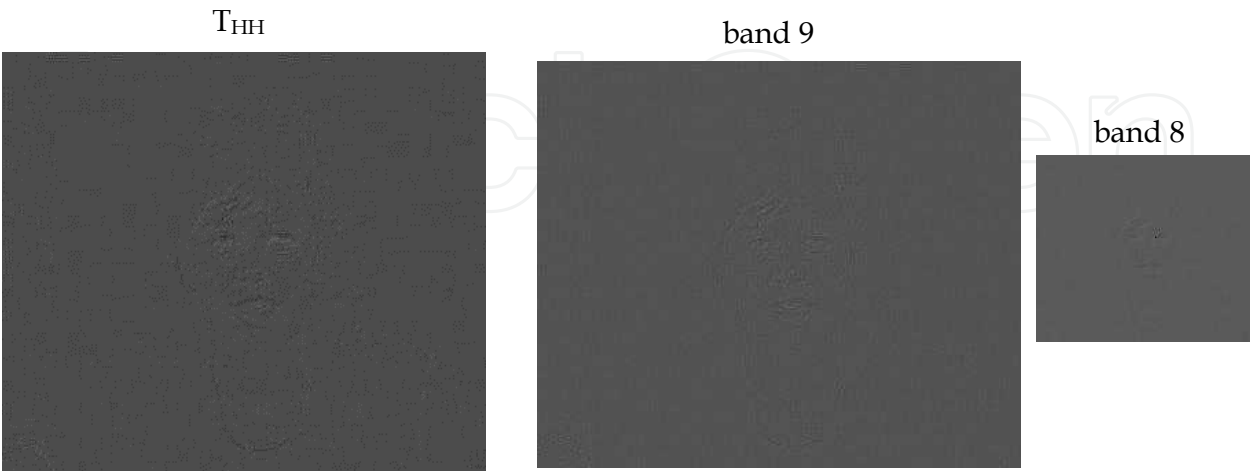

Fig. 8. Spatial subands for the temporal high-high band.

5. Local adaptive vector quantization

Once the original image has been decomposed and the redundancy in the data removed, the next step in the image compression problem is to encode the constituent bands. It has been found that most of the energy signal resides in the lower spatial frequency subands, namely bands 1, 2, and 3. Subbands 4, 5, 6, 7, 8, and 9, which corresponding to the high frequencies, carry most of the motion and edge information and acts as a motion detector. Thus, by accurate coding of low spatial-temporal bands, the spatial details of the original image are conserved. Unlike the majority of the works on the structuring of VQ codebooks, the primary goal of this work is to make the codebook simple and robust to the motions which occur in video sequences and which are seldom capture from a single training sequence. Therefore, the local adaptive vector quantization (LAVQ) [14] is adapted to encode some of the spatial subbands.

A simple and effective one-pass image compression algorithm is provided by the local adaptive vector quantization (LAVQ) [15]. The encoder has a codebook containing codewords (vectors). Each of these codewords is assigned an index corresponding to its position in the codebook. The image is scanned block by block. Each time, a block is taken and compared to the stored codewords. If there is a codeword sufficiently close to the image block (within the pre-decided error) the index itself is sent, and that codeword is moved to the top of the codebook in both transmitter and receiver. If the codebook search is complete without accepted codeword, a special index is sent and followed by the block itself. Now, this block is considered a new codeword and is placed at the top of the codebook. All other codewords are pushed down, and if the number of codewords exceeds the maximum allowed, the last codeword is lost [14]. The LAVQ algorithm maintains the most recently used vectors in the codebook in order of last usage. This allows the LAVQ algorithm to be quick and efficient for any image to be encoded without codebook training. New codewords are generated more often in regions containing edges and fine features, while blank regions are coded with fewer new codewords. Therefore, LAVQ is suitable for the high bands where the correlation is low. The properties of one-pass and high speed codebook generation and encoding are the two main advantages of LAVQ[14].

5.1 Adaptive dead zone for the high subbands

Before we apply the LAVQ on the spatial subbands, a dead zone for each subband based on the number of occurrence of the pixels around zero value is selected. Then, the band is divided into vectors. The number of zero vectors after applying the dead zone will be increased.

So, instead of encoding and sending all vectors, we just encode and send the non-zero vectors with its location. Figure 9 shows an example for band 5 before and after applying the dead zone concept. This process reduces the number of vectors to be transmitted to 20% – 30% of that needed if the dead zone approach is not adopted.

---

[14] Analysis of Coding and Compression strategies for Data Storage and Transmission, Ph. D. thesis, california Instiute of Technology (Sayano, 1992).

[15] Image compression scheme using improved basic-LAVQ and optimized VLC, J. King Saud Univ., Vol. 8, Eng. Sci. (2), pp. 251-266 (Al-Asmari et al., 1996).
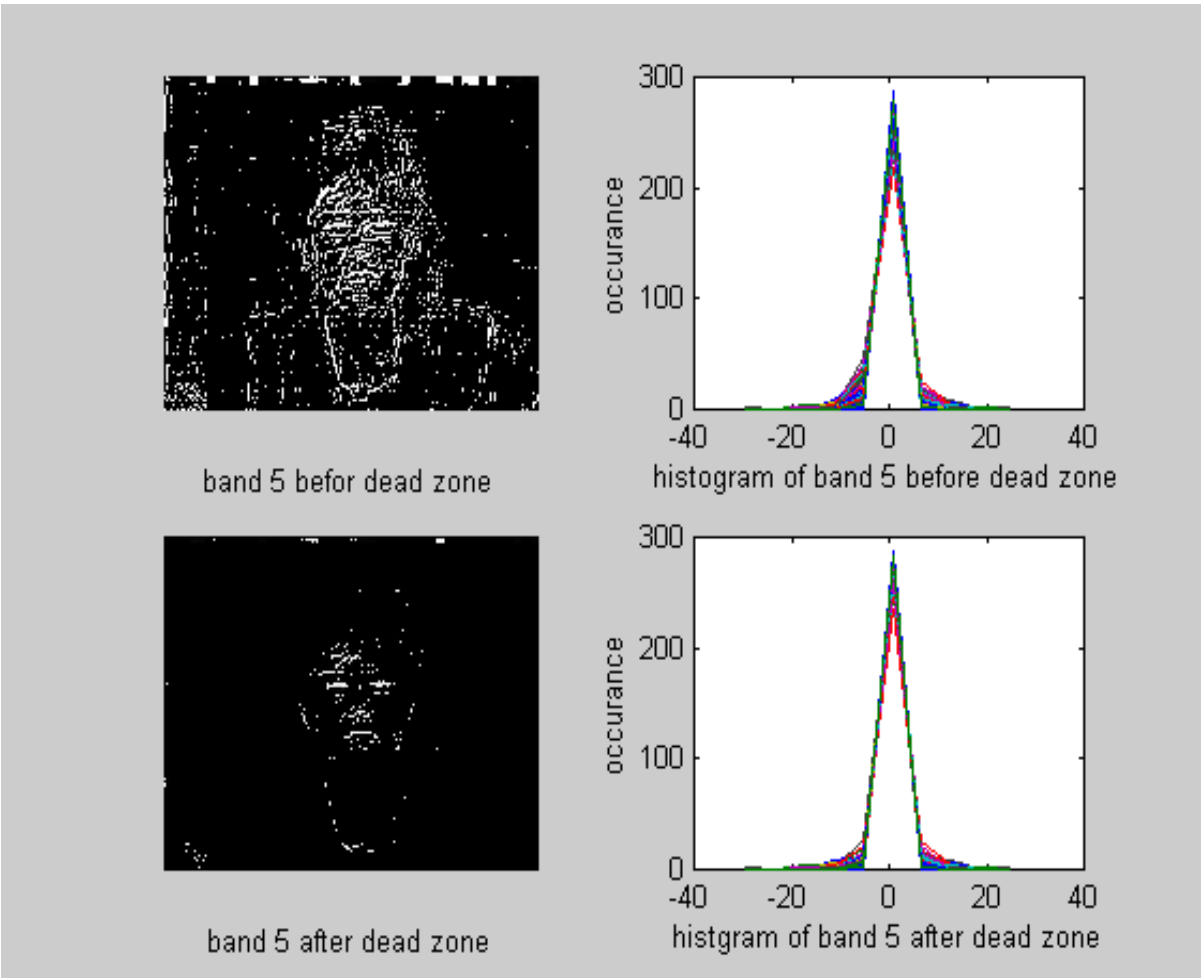
Fig. 9. Example of dead zone process for band 5.

5.2 Searching method for the LAVQ

The difference (error) between the image vectors and the codebook vectors can be calculated using LAVQ searching method concept as follow:

$$d_m = \left(\sum_{j=1}^{cd} \frac{\left(V_i(j) - V_m(j)\right)^2}{cd} \right)^{1/2} \quad (m = 1, 2, ... C_s \text{ and } i = 1, 1 + C_d, 1 + 2C_d..., mxn) \tag{1}$$

Where

d_m is the root mean square error (RMSE).

V_i is the image vector.

V_m is the codebook vector.

c_s is the codebook size.

c_d is the codeword dimension.

The RMSE (d_m) will be compared with a pre-decided threshold error (V_{th}). If d_m is less than this threshold, then the index of the codebook vector with lowest error will be transmitted and this vector is moved to the top of the codebook in both transmitter and receiver. Otherwise, the image vector will be transmitted and placed at the top of the codebook.

6. Encoding of the subbands

High bit allocation is assigned for the baseband (band 1) since most of the energy of the decomposed T_{LL} image is concentrated in this band. For the high bands (band 2 – band 9), different encoding algorithms are design to be tested for these bands. The first encoding algorithm is to test the LAVQ for all the high bands (band 2 – band 9). The second technique is to encode some of these bands using the edge detection concept, then applay the LAVQ on the detected information. The third approach is to encode some of these high bands using the absolute moment block truncation coding (AMBTC) algorithm[16]. The overall encoding algorithm is decided based on the highest performance in terms of the peak signal to noise ration (PSNR) and the visual quality for the encoded subband. For the first algorithm, those bands reconstructed with PSNR greater than 51 dB will be encoded with LAVQ. The second algorithm is adapted for those subbands encoded with PSNR greater than 40 dB and with bit rate less than 0.06 bpp. The overall encoding algorithm is decided based on the best performance for each subband.

The second level in the pyramid has the high frequency content of the decomposed temporal T_{LL} subband. From the simulation results of the previous three algorithms,
it is found that AMBTC algorithm will give the best performance for this subband (band 2) because it is highly correlated than the other subbands. The mean, absolute moment and bit map are transmitted for each subblock.

The first difference level (band 3) in the pyramid has the minimum information most of which is concentrated around the edges. This information is encoded by applying an edge detection approach to find the location of pixels that are perceptually important and then transmitting only these encoded pixels. To avoid the transmission of the position of the encoded locations, a predicting scheme for the edges of band 3 from the encoded bands (band 1 and band 2) is adapted at the receiver side. This is shown in Figure 10.
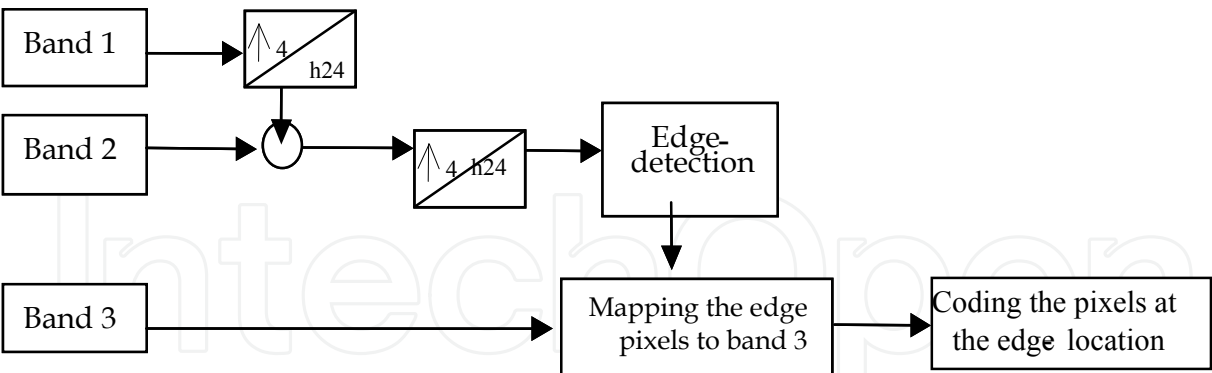

Fig. 10. Coding of Edge-detection for band 3.

Before encoding, the baseband is interpolated to the second level and added to band 2. Then, the result is further interpolated to the size of band 3. Edge-detection is applied to this up-sampled version and the corresponding pixels from band 3 are formed into vectors to be encoded for transmission. At the receiver, similar process is repeated by upsampling the encoded baseband to be added to the encoded band 2. Then, these two subbands are up-

---

[16] Absolute moment block truncation coding and application to color images, IEEE Trans. on Commun., Vol. COM-32, pp-1148-1157 (Lemma & Mitchell, Oct. 1984).

sampled to have the same size as band 3. This partially reconstructed version of the original image is used for edge-detection, which gives the location of the vectors of the band 3 that are encoded. Once the locations are known, the encoded vectors are suitably placed to form band 3. Thus, using this approach, no side information needs to be sent for the encoded areas of the first difference image level and an average of only 4 % to 5% of this level needs to be encoded. Thus, more compression is achieved by edge-detection instead of coding the entire band. After edge-detection, the data is encoded by using LAVQ.

The decomposed subbands (band 4 and band 5) of the temporal (T_{LH}) subband are encoded using the LAVQ for band 4 and the edge-detection approach for band 5. Band 4 is interpolated to the size of band 5. Then, the edge-detection technique is applied to the interpolated version and the corresponding pixels from band 5 are formed into vectors to be transmitted. At the receiver, this process is repeated by interpolating the decoded band 4 to the size of band 5. Then, edge-detection is taken for this band to decide the location of the received vectors for band 5. The same decoding process is adapted for bands 6 and 7 to get the temporal T_{HL} subband. Band 8 and band 9 formed the temporal T_{HH} subband. Band 9 has extremely low energy content and the sparse information carried in this band is not significant for the final image reconstruction. Thus, it can be safely discarded. Band 8 is encoded using the LAVQ algorithm. At the receiver, this band is interpolated to get the temporal T_{HH} subband.

The subbands encoded using LAVQ have different codebooks sizes and vectors (codewords) dimension. The choice of the codebook size and the codeword dimension depends on many factors such as the important of the information to be transmitted, the correlation between the data in each band, and the size of the band to be encoded. Band 4, for example, is more important than band 8 because it has some of the motion information, and also used at the receiver as a detector to encode band 5. Therefore, some care shall be given to this band. The codebook size is selected to be 128 codewords each of dimension 4. Band 6 and band 8 have most of the edges information, which already emphasized by encoding band 7. Therefore, a codebook with smaller size and a codeword of bigger dimension than that of band 4 can be adapted. From the simulation results, we find that the codebook of size 64 with codeword dimension of 8 will give an excellent reconstructed video sequence quality.

Since the encoded vectors corresponding to the edge-detection concept are the only information used to reconstruct band 3, band 5, and band 7, then, these vectors shall be quantized with lowest possible MSE. Therefore, codebooks of size 64 and codewords of dimension 4 pixels are selected to encode those subbands. Table 3 shows the encoding techniques, PSNR, and bit rate for Miss America sequence.

	Encoding technique	PSNR	Avg. Bit rate (bpp)
Band 1	AMBTC	45	0.0027
Band 2	AMBTC	43.37	0.02536
Band 3	LAVQ + Edge-detection	40.3749	0.0543
Band 4	LAVQ	52.6049	0.0133
Band 5	LAVQ + Edge-detection	47.1512	0.0061
Band 6	LAVQ	54.2208	0.0123
Band 7	LAVQ + Edge-detection	40.9862	0.0086
Band 8	LAVQ	51.7706	0.0145
Average		37.2	**0.137**

Table 3. The average bit rates and PSNR for Miss America sequence with V_{th} = 9.

7. Simulation results

The compression algorithm is tested on three video sequences with different motions and backgrounds. The first sequence is Miss America sequence with slow motion and static background. In this sequence, the only moving objects are the lips and head. The second sequence with moderate motion and noisy background is the Salesman sequence. The man's head and hand are moving faster than Miss America sequence and with noisy background. The third sequence with fast motion than the man's sequence is Walter sequence. All of these sequences were 256 level gray-scale images with dimension of 288 x 352 pixels per frame at the rate of 30 frames /s and 8 bits per pixel. These sequences are standard and are used by many researchers. The only way to check the performance of our proposed algorithm is to test it on such images and compare the results with other compression algorithms.

7.1 Calculating the bit rates

The baseband (band 1 with 18×22 pixels) and band 2 (with 72×88 pixels) are encoded using AMBTC because the information in those subbands is highly correlated.
Band 1 and band 2 are divided into sub-blocks to be encoded by using AMBTC algorithm. The average bit rate needed to encode each band is calculated as follow:

$$Bitrate = \frac{(B_m + B_{h-l}) \times (\dfrac{\frac{l_1}{4^n} \times \frac{l_2}{4^n}}{B_d})}{l_1 \times l_2} \tag{2}$$

Where B_m is the bit-map, B_{h-1} is the required bits to encode the high and the low mean, B_d is the sub-block dimension, $l_1 \times l_2$ is the original image frame dimension, and n is the pyramid level number (n = 1, for first level, and n = 2 for second level). In this paper, B_d is selected to be 3x3 for band 1, and 4x4 for band 2. The high mean and the low mean for band 1 are encoded at 8 bits each, while for band 2, the high mean and the low mean are encoded at 6 and 4 bits; respectively.
The bit rate calculation for the Miss America sequence according to equation (2) is shown below for four frames of the original image sequence; namely, frame 4,5,6 and frame 7. These frames are considered to have the highest motion among this test sequence.

$$Bitrate(band1) = \frac{(9+16) \times \dfrac{18 \times 22}{3 \times 3}}{288 \times 352} = \frac{0.01085}{4} = 0.0027 bpp \tag{3}$$

$$Bitrate(band2) = \frac{(16+10) \times \dfrac{72 \times 88}{4 \times 4}}{288 \times 352} = \frac{0.10156}{4} = 0.0253 bpp \tag{4}$$

The second encoding technique (edge detection + LAVQ) is used to encode bands 3, 5, and 7. First, the edge concept is applied for those bands. Then, those pixels corresponding to the positions represented with "1" are encoded using the LAVQ technique.
The first algorithm (LAVQ) is found to be of high performance for bands 4,6 and 8 since the motion and edges information are presented in these bands. The average bit rate for each band can be calculated using the following formula:

$$Bitrate = \frac{(X \times b) + (Y \times ((cd \times 8) + b))}{l_1 \times l_2} \tag{5}$$

Where X is the number of matched codewords, b is the number of bits needed to encode the index, Y is the number of non-matched codewords, and cd is the dimension of codeword.

Since the eight subbands represent four frames, then the bit rate for each frame is given by the sum of the bit rates for each band divided by a factor of four.

Simulation results on these sequences can be discussed based on three main factors: peak signal to noise ration (PSNR), bit per pixel (bpp) and visual quality. The PSNR in decibel (dB) is given by;

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \tag{6}$$

Where MSE is the mean square error written as follow;

$$MSE = \frac{1}{mxn}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij} - \tilde{X}_{ij})^2 \tag{7}$$

where;

m is the number of rows.

n is the number of columns.

$\tilde{X}$ is the reconstructed pixel value.

Table 4 shows the average bit rate and average PSNR with different pre-decided thresholds error (V_{th}) for the three sequences. These thresholds are compared with the RMSE (dm) that result from LAVQ searching method. The bigger the V_{th} the lower number of non-matched (Y) codewords will be. This will reduce the average bit rate required for transmission as given in equation (5). However, the quality of the reconstructed image will be effected. Therefore, a compromised between the bit rate and the required visual quality shall be decided.

		Miss America	Salesman	Walter
	PSNR	37.34	36.5	35.6
V_{th} = 6	**Bpp required**	0.143	0.172	0.207
	Bit rate	0.435 Mbps	0.523 Mbps	0.629 Mbps
	PSNR	37.2	36.31	35.42
V_{th} = 9	**Bpp required**	0.137	0.161	0.198
	Bit rate	0.416 Mbps	0.489 Mbps	0.602 Mbps
	PSNR	36.9	36.2	35.15
V_{th} = 12	**Bpp required**	0.13	0.154	0.191
	Bit rate	0.395 Mbps	0.468 Mbps	0.58 Mbps

Table 4. Performance characteristics for different test sequences.

Three different thresholds have been tested in this study. The best result in terms of PSNR (dB) is given at V_{th} = 6. However, at this threshold the bit rate requirement is higher than that at $V_{th} > 6$. At V_{th} = 12, the bit rate for the three sequences is very low. It has been found from the simulation results that the visual quality of the reconstructed sequences is excellent

at V_{th} = 12. Figure 11 shows the original and the reconstruction of the second four frames of Miss America sequence at V_{th} =12. The visual quality of the reconstructed sequence is the same as the original. From the simulation, it can be concluded that this compression algorithm is capable of compressing a video sequences of different motions.
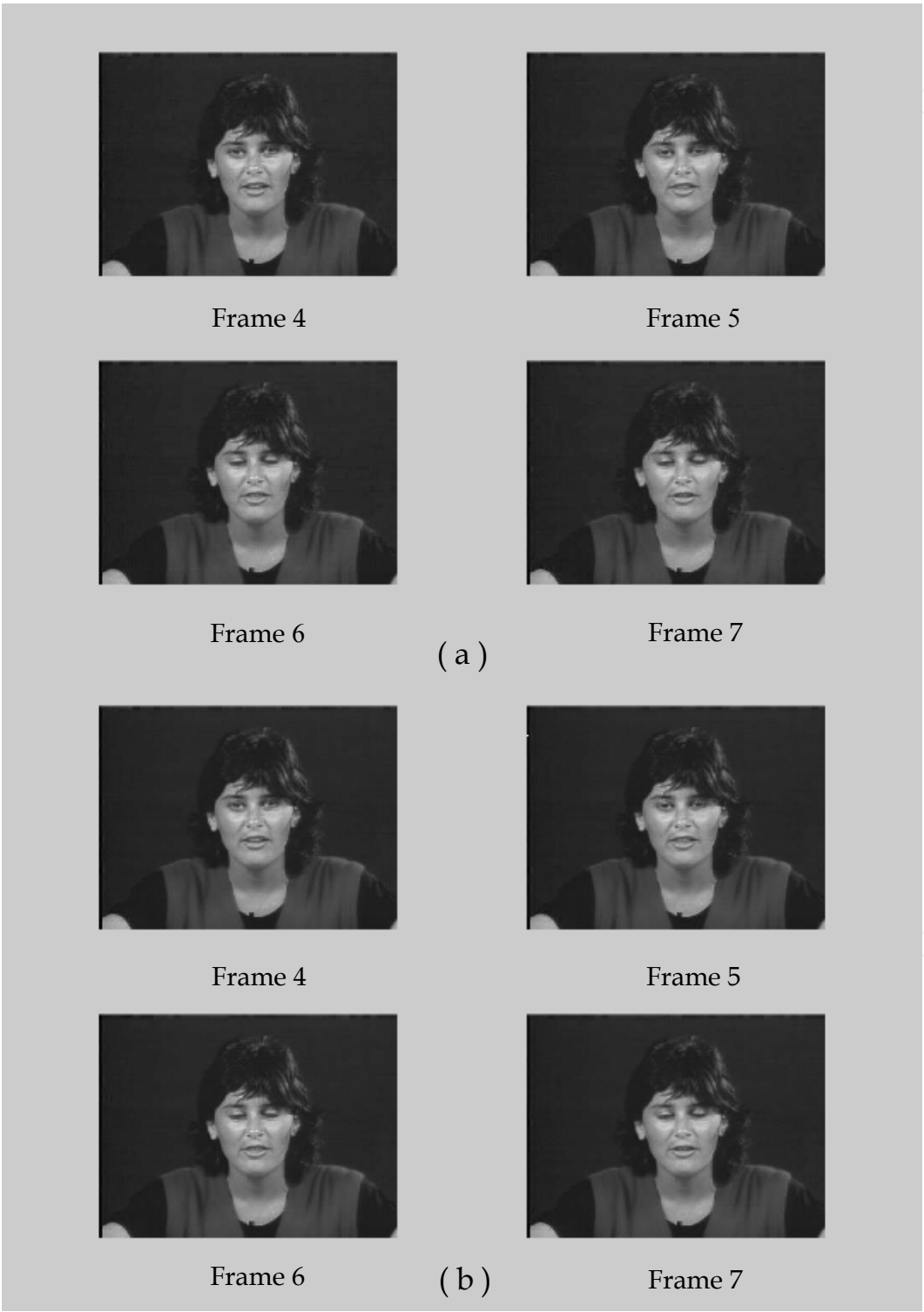

Fig. 11. (a) Original and (b) Reconstructed frames of the second four frames of Miss America sequence at V_{th} = 12.

The PSNR (dB) via the number of frames is demonstrated for the three sequences at different thresholds. Also the average bit rate (bpp) versus the number of frames is presented for 16 frames of the test video sequence. Since four frames are simulated as one group at a time, only 4 different bit rates will be observed. However, the subband T_{LL} is transmitted first. Then, band T_{LH} which is the second important subband regarding the information contents. This approach will be followed until all the subbands are transmitted. Accumulative calculation for the bit rate is then adopted in order to plot the bit rate curves. Figure 12 shows the overall performance in terms of PSNR (dB) and bit rate (bpp) of the proposed algorithm for Miss America sequence.
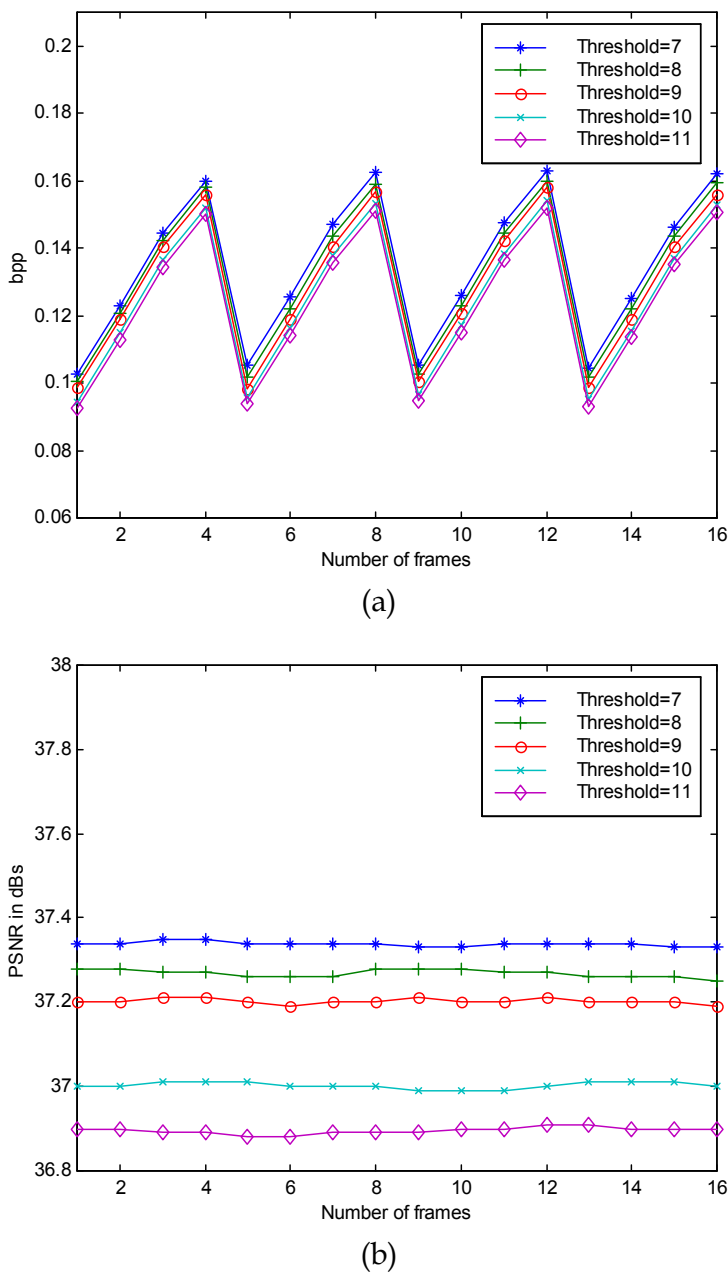

(a)

(b)

Fig. 12. Performance curves. (a) PSNR vs. number of frames for Miss America sequence. (b) bpp vs. number of frames for Miss America sequence.

8. Performance evaluation

The performance evaluation of the proposed algorithm is done in two stages. First, it is compared with the performance of MPEG standard algorithm. Second, it is compared with some existing research works in this field using multiresolution decomposition concept.

8.1 Comparison with MPEG-1

In[17], we have tested MPEG-1 for the monochrome Miss America sequence at 30 frames/second. The results are produced for CIF (288 × 352) image format. The PSNR is 37 dB and the bit rate is 0.343 bpp[17]. The proposed algorithm in this paper achieves 36.9 dB at 0.13 bpp. At approximately the same PSNR, the saving in bit rate is about 56% as shown in Table 5. The visual quality is excellent in both MPEG-1 and the proposed technique. However, the encoder/decoder of MPEG-1 is more complex than our scheme. MPEG-1 based on a coder uses the DCT transform for the intraframe and the motion compensation (MC) for the interframe. Studies reveal that if there are many moving objects in the image, each moving in a different direction, the search technique becomes computationally complex, involving larger storage and delay problems[17].

	MPEG-1	KARL [7]	NEHAL [8]	Al-Asmari [10]	Al-Asmari [17]	This Method
Interframe Coding	Motion compensation	Temporal 2-tap filtering SSKFs	Temporal 2-tap filtering SSKFs	Temporal 2-tap filtering SSKFs	Motion compensation	Temporal 4-tap filtering SSKFs
Intraframe Coding	DCT – Transform (JPEG)	SBC + ADPCM and PCM	PC + DPCM or BTC	PC + FSCL (VQ)	AMBTC and quantizetion	PC + AMBTC + LAVQ
Average PSNR	37 dB	36.9 dB	36.5 dB	36.52 dB	37 dB	36.9 dB
Bpp	0.343	0.434	0.273	0.25	0.2	0.13
Relative complexity	Complex	Moderate	Low	Moderate	Moderate	Low
Comments	Excellent quality with moderate bit rate	Good Quality at high bit rate	Competitive quality with low bit rate	Excellent quality with low bit rate	Excellent quality at Low bit rate	Excellent quality with very low bit rate

Table 5. Comparison with different compression methods.

8.2 Comparing with other works

The performance of the proposed algorithm is compared with the results obtained in related works in[7-8], [10], and[17] in terms of PSNR, bit rate, and complexity. The comparison between those algorithms is presented in Table 5. Karlsson and Vetterli in[7], and Nehal, et al. in[8] have suggested schemes for progressive transmission using temporal filtering with 2-tap symmetric short kernel filters (SSKFs). The reconstruction quality is very good. For[7], the PSNR is 36.9 dB with an average bit rate of 0.434 bpp while for[8] the PSNR is 36.5 dB and the bit rate is 0.273 bpp. Al-Asmari, et al. in [10] have suggested an algorithm for video compression. The 2-tap SSKFs filters are used for temporal domain and pyramid coding for intraframe. The decomposed bands are encoded using VQ called frequency selective competitive learning (FSCL). This technique used the neural network concept to design the

---

[17] Low Complexity Video Compression Algorithm Using AMBTC, Proceeding of IEEE Military communication conference, Atlantic city, NJ (Al-Asmari et al., 31 Oct. – 3 Nov. 1999).

codebook for the vector quantization. This algorithm gives an excellent image quality for the Miss America sequence at an average bit rate of 0.25 bpp and PSNR 36.52 dB. This algorithm is considered to be of higher complexity than our algorithm because of the codebook design. In [17], the authors present a coder based on a combination of AMBTC for intraframe and MC for interframe. They produce results for the monochrome Miss America sequence. For the CIF format (i.e. 288 × 352) at 30 frames/sec, they achieve approximately 37 dB at 0.2 bpp. The disadvantage of this algorithm is the use of motion compensation for fast motion video sequence. For the same video sequence, the proposed algorithm in this paper gives a higher PSNR (37.2 dB) and a lower bit rate (0.13 bpp) than those algorithms for a coder with lower complexity.

9. Conclusion

The results presented here are better than other coding schemes, which are published using almost the same coding technique concept. The 3-D decomposition does not make unrealistic assumptions about the data, as do methods based on motion compensation (MC). Moreover, coding and decoding for the proposed algorithm are of comparable and relatively low complexity. The robustness obtained by adapting the LAVQ for the codebook has been discussed. The results reported in this paper are independent of which sequence is used to produce the codebook.

This scheme is faster than MPEG algorithms and other existing technique based their encoder on VQ concept since no need for training set or codebook generation. This scheme will be an optimal choice for real time transmission. It is well suited for progressive transmission of the video sequence and for browsing moving images via the Internet.

LAVQ technique gives good performance with those bands, which are highly uncorrelated, and with spark information. Bit rate is varying from 0.13 to 0.191 bpp depending on the nature of the sequence. Different video sequences have been tested and show very good image quality with PSNR in the range of 36.9 to 35.15 dB and with bit rate range from 0.395 Mbps to 0.58 Mbps as shown in table 4.

10. References

Al-asmari Awad Kh. (1995), *Optimum Bit Rate Pyramid Coding with Low Computational and Memory Requirements*, IEEE Trans. Circuit and Systems for video Tech., vol. 5, no. 3 pp. 182-192, (June 1995).

Al-Asmari Awad Kh., Ahmed Abobakr & Al-Doweesh Abdullah (1996), *Image compression scheme using improved basic-LAVQ and optimized VLC, J.* King Saud Univ., Vol. 8, Eng. Sci. (2), pp. 251-266.

Al-Asmari Awad Kh., Aryai Deepali, & Kwatra Subhash C. (2000), *Video Signal Transmission for IS-95* Environment, Electronic Letters, Vol. 36, no. 5, pp. 465-466, 2nd March 2000.

Al-Asmari Awad Kh., Dave Sameep & Kawatra Subhash C. (1999), *Low Complexity Video Compression Algorithm Using AMBTC*, Proceeding of IEEE Military communication conference, Atlantic city, NJ, (31 Oct. – 3 Nov).

Al-Asmari Awad, Singh Vinay & Kawatra Subhash (June 28 – July 1 CIST 99), *Hybrid coding of images for progressive transmission over a digital cellular channel*, CISST'99

International Conference on Imaging Science, Systems and Technology, Monte Carlo resort, Las Vegas, Nevada, USA, PIN 128C.

Bhaskaran Vasudev, Konstantinidides Konstantions & Hewlett Packard Laboratories (1996), *Image and video compression standards: algorithms and architectures*, Kluwer Academic Publishers.

Burt P. & Adelson E. (1983), *The Laplacian Pyramid as a compact Image Code*, IEEE Trans Communication., vol. COM-31, no. 4 pp. 532-540, (April 1983).

Child J. (July 1993), *Realtime video compression poses challenge to designers and vendors alike*, Computer Design, vol. 32, no. 7, pp. 67-70.

Clarke.R. J., (1995) *Digital compression of still images and video*, Academic Press.

Coppisetti N., Kwatra S. C., & Al-asmari A. Kh. (1993), *Low complexity subband encoding for HDTV images*, IEEE J. Select. Areas Commun., vol. 11, no. 1, pp. 77-87, (Jan. 1993).

Dantwala Nehal, Kwatra Subhash C. & Al-Asmari Awad Kh. (June 30 – July 3 - CISST'97), *Pyramid coding of video signals for progressive transmission*, CISST'97 International Conference, pp-392-398.

Karlsson .G. & Vetterli. M. (July 1988), *Subband coding of video for packet networks*, Optical Engineering, vol. 27, no. 7, pp. 574-586.

Lemma Maximo D. & Mitchell O. R. (1984), *Absolute moment block truncation coding and application to color images*, IEEE Trans. on Commun., Vol. COM-32, pp-1148-1157, (Oct. 1984).

Netravali. Arun . N & Haskell Barry G. (1995), *Digital pictures representation, compression, and standard*, Plenium Press.

Sayano M. (1992), *Analysis of Coding and Compression strategies for Data Storage and Transmission*, Ph. D. thesis, california Instiute of Technology.

Sayood Khalid (1996), *Introduction to data compression*, Morgan Kaufmann Publishers Inc., San Francisco, California.

Shannon.C. E (Oct. 1948), *A mathematical theory of communication*, Bell System Tech. J.. Vol. 27, pp. 379-423, July. 1948 and pp. 623-656.

Discrete Wavelet Transforms - Algorithms and Applications

Edited by Prof. Hannu Olkkonen

ISBN 978-953-307-482-5

Hard cover, 296 pages

Publisher InTech

Published online 29, August, 2011

Published in print edition August, 2011

The discrete wavelet transform (DWT) algorithms have a firm position in processing of signals in several areas of research and industry. As DWT provides both octave-scale frequency and spatial timing of the analyzed signal, it is constantly used to solve and treat more and more advanced problems. The present book: Discrete Wavelet Transforms: Algorithms and Applications reviews the recent progress in discrete wavelet transform algorithms and applications. The book covers a wide range of methods (e.g. lifting, shift invariance, multi-scale analysis) for constructing DWTs. The book chapters are organized into four major parts. Part I describes the progress in hardware implementations of the DWT algorithms. Applications include multitone modulation for ADSL and equalization techniques, a scalable architecture for FPGA-implementation, lifting based algorithm for VLSI implementation, comparison between DWT and FFT based OFDM and modified SPIHT codec. Part II addresses image processing algorithms such as multiresolution approach for edge detection, low bit rate image compression, low complexity implementation of CQF wavelets and compression of multi-component images. Part III focuses watermaking DWT algorithms. Finally, Part IV describes shift invariant DWTs, DC lossless property, DWT based analysis and estimation of colored noise and an application of the wavelet Galerkin method. The chapters of the present book consist of both tutorial and highly advanced material. Therefore, the book is intended to be a reference text for graduate students and researchers to obtain state-of-the-art knowledge on specific applications.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Awad Kh. Al-Asmari (2011). Low Bit Rate Video Compression Algorithm Using 3-D Discrete Wavelet Decomposition, Discrete Wavelet Transforms - Algorithms and Applications, Prof. Hannu Olkkonen (Ed.), ISBN: 978-953-307-482-5, InTech, Available from: http://www.intechopen.com/books/discrete-wavelet-transforms-algorithms-and-applications/low-bit-rate-video-compression-algorithm-using-3-d-discrete-wavelet-decomposition