Very Low Bit-Rate Digital Video Coding[1]

LEE SCARGALL and SATNAM DLAY
Department of Electrical and Electronic Engineering
University of Newcastle upon Tyne, Newcastle, NE1 7RU, England

Tel: +44 191 222 8356, Fax: +44 191 222 8180

Abstract: - In this paper, an image sequence coding scheme for very low bit-rate video coding is presented. The new technique utilises windowed overlapped block matching motion compensation for the temporal coding scheme, and vector quantization to reduce the spatial redundancy within the predicted image. When the codec is configured to operate at 11.4 kbit/s, average PSNR values of 33.68dB and 26.71dB are achieved for the "Miss America" and "Carphone" sequences respectively. We propose a new methodology for adaptive vector quantization (AVQ), where the codebook is updated with new vectors. The new vectors replace less significant ones in the codebook based on a novel scoring criterion that utilises a forgetting factor and codebook half-life. The proposed method gives rise to an additional performance enhancement of around 1dB over conventional techniques of AVQ. Furthermore, the methods do not suffer from blocking effects due to the inherent properties of both the temporal and spatial coding. IMACS/IEEE CSCC'99 Proceedings, Pages:5691-5697

Key words: Video coding, Windowed overlapped motion compensation, Adaptive Vector Quantisation.

1. Introduction

Over the past decade there has been significant progress in the field of very low bit-rate image coding schemes for video communications. Various image-coding techniques such as the discrete cosine transform (DCT), wavelet transform and vector quantization (VQ) have all been implemented in various forms. Despite of all the marginal improvements of one technique over the other, the DCT-based codecs have remained to be the favourite within the image/video community.

The international standar dcodecs such as H.261, H.263 and MPEG are all based on the DCT method. Despite the efficiency of the DCT in terms of the energy compaction, when it is employed to remove the spatial redundancy, akin to the way it is implemented in the standards mentioned above [1-3]. The DCT does not lend itself to be an efficient coding scheme, in terms of the achievable coding gain. Before processing, the image is broken into distinct blocks, which have to be small to represent the local changes accurately. These blocks are then processed independently from each other and the interblock redundancy remains unutilised. There are also problems associated with the discrete wavelet transform [4] when it is used in combination with block matching motion compensation (BMMC). Unfortunately, when using this temporal coding scheme, the predicted frames contain sharp edges due to the motion compensated block boundaries. When the wavelet transform is applied to the difference frame, the sharp edges of the block boundaries have a detrimental effect on the achievable coding gain by creating large transform coefficients in the high frequency bands. Hence the importance of an efficient motion compensation scheme in relation to the achievable coding gain in very low bit-rate systems.

It is to some extent believed that future video codecs for very low bit-rates, i.e., below 20 kbit/s, cannot be immune from VQ due to its overwhelming ability to reduce the spatial redundancy [10]. There are many other advantages associated with VQ, especially its capacity to operate in error-resilient coding systems [7]. The main advantages stem from the use of fixed length code words that limit the error propagation to the isolated code words. The impact of the expected channel errors can be minimised even further by mapping the code-vectors onto code words, a process called enumeration. In addition, VQ does not suffer from blocking effects

[1] This paper is complemented by a demonstration package portraying video sequences at various bit-rates, which is down loadable from http://www.students.ncl.ac.uk/l.d.scargall/.

that are visually disjointed and has therefore a major advantage over DCT based methods.

In our endeavours to design a very low bit-rate VQ codec, we examine the performance of various codebooks to remove the spatial redundancy within the difference frame. We also investigate the performance of the proposed image-coding scheme when the codebook of the vector quantiser is updated with new vectors, to take into account the time-variant nature of video sequences. New vectors replace less significant ones in the codebook based on a novel scoring criterion that utilises a forgetting factor to calculate the local importance of each vector.

The organisation of this paper is as follows. The temporal-coding scheme is discussed in Section 2, while the utilisation of the forgetting factor in adaptive VQ schemes is described in Section 3. The bit allocation strategy is outlined in Section 4, while the performances of the image coding schemes are compared to the H.263 codec in Section 5. Finally, some concluding remarks are formulated in Section 6.

2. Motion Compensation

The most widely used technique to exploit temporal redundancy of video signals is motion compensation prediction. In this method, it assumes that parts of the current picture can be modelled as a translation of the previous frame. In conventional BMMC schemes, the source frame is segmented into distinct blocks and the motion vector (MV) for each block is obtained using block matching methods on the previous frame. The predicted current frame is then obtained from the motion compensated previous frame. Unfortunately, this technique causes visible block boundaries to appear in the predicted image and so the achievable coding gain decreases. The image-coding algorithm presented in this paper utilises the windowed overlapped BMMC scheme [5, 14]. In this technique, when the blocks overlap, before merging them into the predicted image, they are weighted by an appropriate window function. Hence, the visible block boundaries are removed from the predicted image and the achievable coding gain increases. The window function used in our temporal coding scheme is the classical raised cosine.

3. Spatial Coding Scheme

The image-coding scheme is suitable for 176 x 144 pixels QCIF videophone sequences, with a scanning rate of 10 frames per second. The codec is configured to operate at fixed bit-rates around 10 kbit/s. The schematic of the proposed scheme is illustrated in reference [15]. The operation of the system is as follows. During the call set up phase a low-resolution intra-frame is transmitted to the decoder to initialise the reconstruction frame buffers. The codec then switches over and remains in the inter-frame mode. Motion compensation is then applied to remove the temporal redundancy within the image. The difference frame is then passed onto the VQ stage to remove the spatial energy.

3.1. Intra-frame Coding

The purpose of the intra-frame is to provide an initial frame for both the encoder and decoder's reconstruction frame buffers. This is accomplished by dividing the QCIF frame into perfectly tiling square blocks and coarsely quantizing the luminance average of each block. It was found appropriate to quantize the block averages to sixteen levels, which are uniformly distributed between the absolute pixel values of [0,…,255]. In order to transmit the intra-frame within the available bit-rate, the block size had to be adjusted accordingly. When the intra-frame block size did not perfectly tile the QCIF frame, the codec increased the block size and padded the QCIF frame along the fringes.

The scheme also makes use of a partial forced update (PFU) to mitigate the effects of transmission errors. This is carried out on both the encoder and decoder's reconstruction frame buffers, on a similar basis to the intra-frame. In each frame, a certain number of 8 x 8 pixel blocks scattered over the entire frame are periodically refreshed using the 7-bit encoded block averages. The PFU blocks are weighted by a factor of 0.3 and superimposed onto the contents of the reconstructed frame buffer, which is appropriately scaled by 0.7.

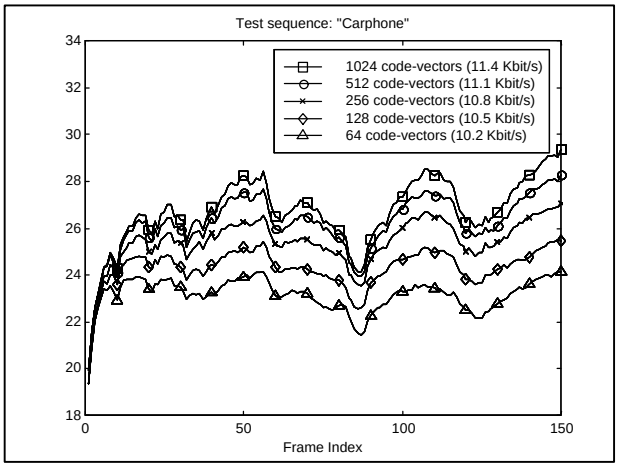
3.2. Vector Quantization

Vector quantization can be used as a powerful means of data compression by comparing an input vector to a finite set of reproduction code-vectors. The comparison measure is based upon minimising the distortion penalty incurred representing a vector with

one of the code-vectors. In theory, VQ can achieve a performance close to the rate distortion bound as the dimension of the code-vectors approach infinity. However, increasing the number of code-vectors in the codebook results in an increase in the encoding complexity due to the codebook search.

To effectively reduce the spatial redundancy in the difference frame, we tested the performance of a VQ codec that contained 1024, 512, 256, 128 and 64 code-vectors in the codebook. In our endeavours to design the codebooks, we generated the MCER frame for various inputs of a QCIF head and shoulders video sequence. In previous research [9], we found that 70% of the spatial energy is contained within 50 out of the 396 8 x 8 pixel blocks. Based on this realisation, the 50 blocks that contained the highest energy in the MCER frame were copied into the training sequence. This ensured that the training sequence of 1500 vectors provided a sufficient variation to generate a statistically independent codebook. The different length codebooks were then generated from the training sequence, using the pairwise nearest neighbour algorithm [10].

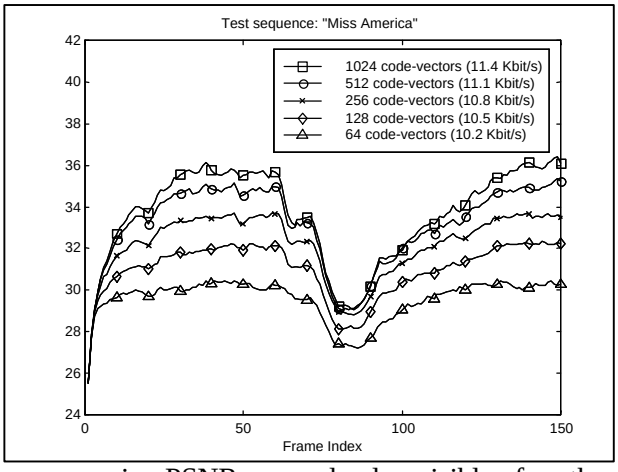
Fig1: PSNR performance of the Carphone sequence.



We tested the performance of the codec by simulating the “Carphone” and “Miss America” video sequences using the defined codebooks. The peak signal-to-noise ratio (PSNR) performance (luminance component only) for the two test sequences are portrayed in Figure 1 and Figure 2 respectively. The average PSNR values are also tabulated for both sequences in Table 1. Although each codec has a different number of vectors in the codebook, all codecs still process the same number of vectors in the temporal-spatial schemes.

Fig 2: PSNR performance of the Miss America sequence.

With reference to Figures 1 and 2, the achievable



increases in PSNR are clearly visible for the different codebooks. Obviously, the larger the codebooks, the better the performance due to the increase in available vectors that can accurately represent the MCER. However, larger codebooks are handicapped by the associated increase in bandwidth to represent each vector in the codebook. When the codec was configured to operate at 11.4 kbit/s, average PSNR values of 26.71dB and 33.68dB were achieved for the “Miss America” and “Carphone” sequences respectively.

3.3. Adaptive Vector Quantization

In AVQ [16], the codebook is typically updated with vectors that occur frequently but are not represented in the codebook. The increased flexibility of this method is unfortunately associated with an inflated bandwidth requirement, to transmit the new vectors to the receiver.

In our adaptive approach, new vectors replace less significant ones in the codebook based upon a novel scoring system. Every time a vector is selected from the codebook, it inherits a unit score. Hence, a vector that is used more frequently is deemed more significant, will have a higher score. However, it is fair to assume that the time-variant nature of video sequences may cause a vector to gain an artificially high score, if it is used excessively at one period and remains redundant for the rest. For this purpose, it is vital to take into account the local scores of each vector. This is accomplished by multiplying the scores of the codebook by a forgetting factor. The forgetting factor a, is a number less than one, and is determined by the pre-selected codebook half-life z. This is mathematically represented by equation

(1). After each frame the codebook scores are multiplied by the forgetting factor to ensure that the local significance of each vector is preserved.

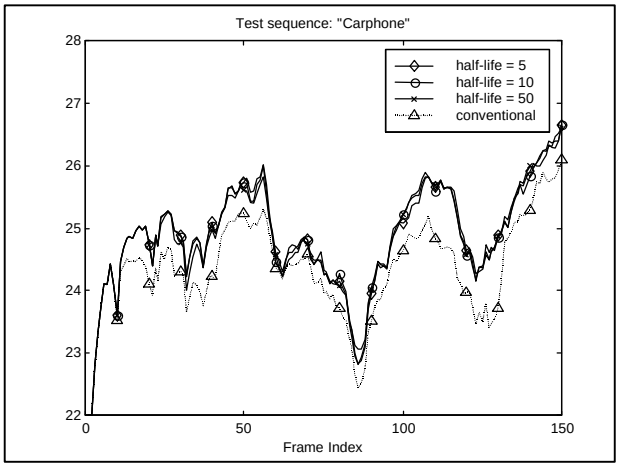
$$(0.5) = a^{z} \qquad (1)$$

When a new vector is initiated into the codebook, it must be given a chance to establish itself. For this reason, the unit score is inappropriate, so the new vector is assigned the mean value M. This is represented by equation (2), where $P(X)$ is the probability that a vector will be selected from the codebook.

$$M = \sum_{i=0}^{\infty} a^{i}.P(X) \qquad (2)$$

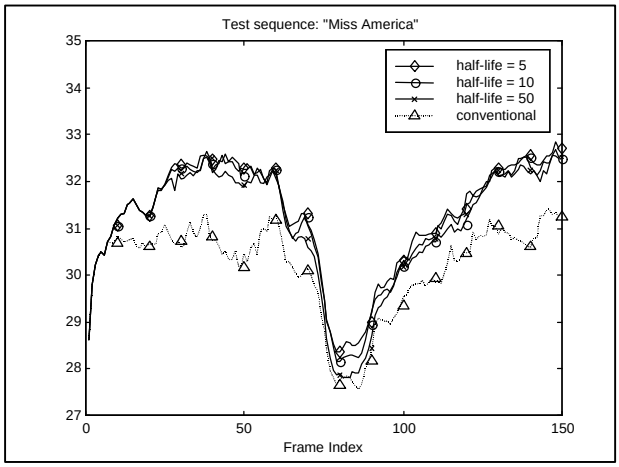
In our endeavours to determine the optimum half-life for a video sequence, we once again simulated the test sequences for half-lives of 5, 10 and 50 frames. These results are illustrated in Figures 3 and 4, and are consistent for a codebook of 64 entries and 2 new vectors per frame. The average PSNR performance for the various half-lives and the performance of the conventional AVQ method is portrayed in Table 2.

Fig 3: PSNR performance of the AVQ(64-2) codec for various half-lives using the Carphone sequence.



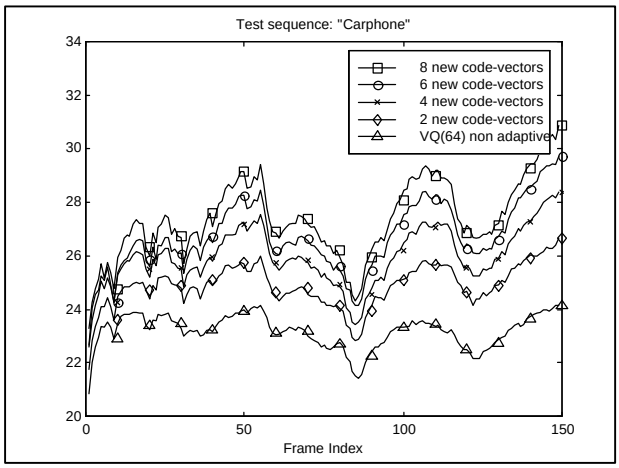
We can impart that the proposed method of AVQ, which utilises the forgetting factor, is superior to conventional methods that rely on frequency scoring alone. With reference to Figures 3 and 4, our results clearly indicate the superiority of the proposed method, which gives rise to an additional PSNR performance enhancement of around 1dB for both test sequences. We can also deduce from the simulation results that the optimum half-life is 5 frames.

Fig 4: PSNR performance of the AVQ(64-2) codec for various half- lives using the Miss America sequence.



We configured the codec to operate in the adaptive mode with a codebook that contained 64 code-vectors and re-simulated the sequences with updates ranging from 2, 4, 8, and 16 new vectors per frame. The PSNR performance is portrayed for both sequences in Figures 5 and 6, with the average values for the whole sequence are tabulated in Table 3.

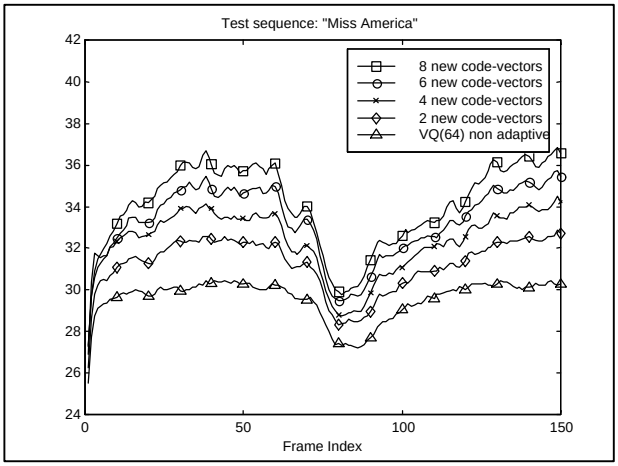
Fig 5: PSNR performance of the AVQ codec with various updates per frame.



Carefully inspecting these findings, we reveal that for every 2 new vectors, the PSNR performance increases by an average of 1dB. We have also shown in Table 3, the additional bandwidth requirement for the various updates. With reference to the statistical properties of the vector training sequence [9], we found that the vectors contained within the MCER are typically distributed between [-8,…,8]. In adaptive VQ, the new vectors in our

system are transmitted to the receiver using a 4-bit linear quantizer bound between these limits. Hence to transmit one new vector, it requires 64 x 4-bit = 256 bits.

Fig 6: PSNR performance of the AVQ codec with various updates per frame.



4. Bit configuration

The bit stream output consists of the frame alignment word (FAW), the PFU data, and the MV and VQ information. The break down of the bit allocation scheme is depicted in Table 4. The 22 bit FAW is required to assist the decoder to regain frame synchronisation after a corrupted frame. The partial forced update refreshes 22 8 x 8 pixel blocks out of the 396 blocks per frame using the 7-bit block averages. Therefore after every 18 frames or 1.8 seconds the update refreshes the same blocks. This periodicity is signalled to the decoder by inverting the FAW.

Previous research [11] has found that the optimum performance for very low bit-rate image codecs, is achieved when the number of MC blocks in the temporal coding scheme is equal to the number of VQ blocks in the spatial coding scheme. Based on this realisation, we configured the image codec to operate on 30 blocks for both the MC and VQ stages. Hence, this implies a video source bit-rate of around 10 kbit/s.

Previous research [9] has also indicated that the optimum search window in terms of the MCER energy reduction is 4 x 4 pixels around the centre of each block. This implies a 4-bit requirement for encoding the 16 possible combinations of the X and Y displacements, and a further 9-bits are required to identify one of the 396 block indices using the enumerative method. Hence, to represent one motion vector it requires a total of 13-bits. For the VQ stage, it too requires 9-bits to identify one of the 396 block indices and a further 2^N bits to represent the codebook index.

5. Comparison of bench mark codecs

As a comparative basis, we configured the ITU-T standard H.263 codec to operate at a target bit-rate of 20 kbit/s and a frame rate of 10 fps. In our PSNR comparison, illustrated in Figure 7, we compare the H.263 codec to our adaptive AVQ(1024-4) codec with 4 updates per frame, and to our VQ(1024) codec that has a bit-rate of 11.4 kbit/s. Simulations were carried out for the "carphone" sequence and the average PSNR values were 28.4dB, 28.0dB and 26.7dB respectively. Hence the performance of the proposed adaptive scheme achieves a comparable PSNR performance to the H.263 codec. However, the delay of our codecs is in principle limited to one frame, while the delay of the H.263 codec stretches to several frames due to the P-frames. This may become a fundamental flaw in the ability of the H.263 codec in offering video communications to people who rely on lip reading or sign language.

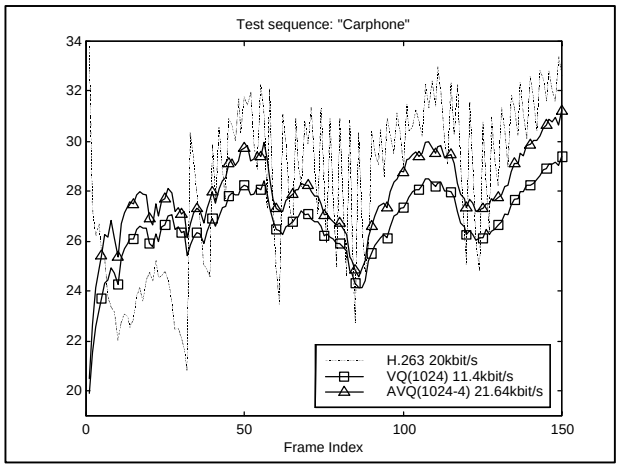


Fig 7: Comparison of the H.263 codec to the proposed VQ schemes for the "Carphone" sequence.

A comparison of the subjective visual quality by way of reconstructed frames from the output sequence was also performed. Unbiased test persons confirmed that the proposed AVQ encoded images are sharper, contain more detail and are without blocking effects.

6. Concluding remarks

The proposed image sequence coding algorithm is a highly efficient coder that is suitable for very low bit-rates. The simplistic nature of the algorithm endears itself even further, offering the best compromise between implementation complexity and performance. The resulting image quality of the VQ codec at 11.4 kbit/s, achieves an average PSNR of 33.68dB and 26.71dB for the "Miss America" and "Carphone" sequences respectively. This is the consequence of the efficient motion compensation scheme in combination with the powerful compression of vector quantization.

A new methodology has also been presented for AVQ, which utilizes a novel-scoring criterion based on a forgetting factor and codebook half-life. The new method gives rise to an additional performance enhancement of around 1dB over conventional techniques of AVQ that rely on frequency scoring alone. The performance of the proposed adaptive scheme achieves a similar PSNR performance to the H.263 standard codec at the same bit-rate. However, the delay of our codecs is in principle limited to one frame, while the delay of the H.263 codec stretches to several frames due to the P-frames. This may become a fundamental flaw in the ability of the H.263 in offering video communications to people who rely on lip reading or sign language.

7. Acknowledgements

The authors would like to gratefully acknowledge the financial support of the Engineering and Physical Science Research Council. Recognition is also given to the Trustees of the Vodafone Charitable Trust, for their benevolent donation. Finally, L Scargall would also like to acknowledge the financial support from the IEE-Vodafone research scholarship.

8. References

[1] ITU-T, *Video Coding for Low Bit-rate Communication*, Recommendation H.263-Draft, July 1995.

[2] T. Sikora: "The MPEG-4 Video Standard Verification Model", IEEE Trans. on C.S.V.T., Vol. 7, No. 1 pp 19-31, Feb. 1997.

[3] A. K. Jain: Fundamentals of Digital Image Processing, Prentice-Hall, 1989.

[4] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 11 pp 674-693, 1989.

[5] S. Nogaki and M. Ohta, "An overlapped block motion compensation for high quality motion picture coding," in Proc. IEEE Int. Symp. Circuits Syst., ISCAS'92, vol. 1, pp 184-187, 1992.

[6] K. Cinkler, "Very Low Bit-Rate Wavelet Video Coding," IEEE Trans. Selected area in Comms., Vol. 16, No. 1, pp 4-11, Jan. 1998.

[7] J. Streit, L. Hanzo: "Dual-Mode Vector Quantised Low Rate Cordless Videophone Systems for Indoors and Outdoors Applications", IEEE Trans. on V.T., Vol. 46, No. 2, pp 340-357, May 1997

[8] R. Stedman, H. Gharavi, L. Hanzo, R.Steele : "Transmission of Subband-coded Images via Mobile Channels", IEEE Tr. on Circuits and Systems for Video Technology; Vol.3, No. 1, pp 15-27, Feb 1993

[9] L. Scargall and S. S. Dlay, "A Comparison of Vector Quantised Video Codecs", Proc. of SPIE-VVD'98, Nov. 1-6, 1998, Boston, MA, America.

[10] A. Gersho, R.M. Gray: "*Vector Quantization and Signal Compression*", Kluwer Academic Publishers, 1992

[11] L. Scargall, S. S. Dlay, "Mobile Videophone Systems for Radio Speech Channels", Proc. of SPIE'98, Jan. 24-30, 1998, California, America, pp 170-179.

[12] J. Streit, L. Hanzo: "Adaptive Low-rate Wireless Videophone Schemes", IEEE Trans. on C.S.V.T., Vol. 5, No. 4, pp 305-318, Aug 1995.

[13] J. Streit, L. Hanzo: "Quadtree-Based Reconfigurable Cordless Videophone Systems", IEEE Trans. on C.S.V.T., Vol. 6, No. 2, pp 225-237, Apr 1996.

[14] J. Katto, J. Ohki, S. Nogaki, M. Ohta: "A Wavelet Codec With Overlapped Motion Compensation for Very Low Bit-Rate Environment", IEEE Trans. on C.S.V.T., Vol. 4, No. 3, pp 328-338, June 1994.

[15] L. Scargall, S. S. Dlay, "Very Low Bit Rate Vector Quantised Video Codecs", SPIE99, Jan. 1999, San Jose, America.

[16] J. E. Fowler: "Generalised Threshold Replenishment: An Adaptive Vector Quantizer Algorithm for the Coding of Nonstationary Sources", IEEE Trans. on Image Processing, vol. 7, no. 10, pp. 1410-1424, Oct 1998.

TABLES OF PERFORMANCE

Code-vectors	Bit-rate in Kbit/s	“Miss America”	“Carphone”
64	10.2	29.58dB	23.16dB
128	10.5	31.01dB	24.31dB
256	10.8	32.19dB	25.42dB
512	11.1	33.15dB	26.20dB
1024	11.4	33.68dB	26.71dB

Table 1: Average PSNR performance of the VQ codec.

Half-life in frames (Codebook = 64, 2 new vectors)	Test Video Sequence	
	“Miss America”	“Carphone”
Conventional method	30.15dB	24.35dB
50	31.03dB	24.83dB
10	31.11dB	24.86dB
5	31.22dB	24.95dB

Table 2: Average PSNR performance for the conventional and half-life codecs.

New vectors per frame (codebook 64, hl = 5)	Additional bit-rate per frame	Test Video Sequence	
		“Miss America”	“Carphone”
0	+0	29.58dB	23.16dB
2	+512	31.22dB	24.95dB
4	+1024	32.37dB	25.93dB
6	+1536	33.26dB	26.65dB
8	+2048	34.09dB	27.34dB

Table 3: Average PSNR values for the AVQ(64) codec with various updates.

Code-vectors	FAW	PFU	MV	VQ	Padding	Bit-rate
64	22	22 x 7	30 x (4 + 9)	30 x (6 + 9)	4	1020
128	22	22 x 7	30 x (4 + 9)	30 x (7 + 9)	4	1050
256	22	22 x 7	30 x (4 + 9)	30 x (8 + 9)	4	1080
512	22	22 x 7	30 x (4 + 9)	30 x (9 + 9)	4	1110
1024	22	22 x 7	30 x (4 + 9)	30 x (10 + 9)	4	1140

Table 4: Bit allocation strategy.