Fixed-Rate Video Codecs for Mobile Radio Systems [1]

JURGEN STREIT, LAJOS HANZO
Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
lh@ecs.soton.ac.uk

Abstract. A comparative study of arbitrarily programmable, but fixed-rate videophone codecs using quarter common intermediate format (QCIF) video sequences scanned at 10 frames/s is offered. In contrast to existing and forthcoming standard schemes, such as the H.261, H.263 and MPEG2, MPEG4 codecs, which rely on bandwidth-efficient but error-sensitive variable-length coding techniques combined with a complex self-descriptive bitstream structure, the proposed codecs exhibit a more robust, regular bitstream and a constant bitrate. Clearly, their philosophy is different from the above error-sensitive and variable-rate standard schemes, since these constant-rate codecs were designed to allow direct replacement of mobile radio voice codecs in second generation wireless systems, such as the Pan-European GSM, the American IS-54 and IS-95 as well as the Japanese systems, operating at 13, 8, 9.6 and 6.7 kbit/s, respectively. This philosophy can, however, be adopted to higher-rate systems, such as the Digital European Cordless Telecommunications (DECT) and the Universal Mobile Telecommunications System (UMTS). The Type 1 codecs proposed benefit from invoking sophisticated compression techniques in order to achieve best video quality at a given bitrate. In contrast, the Type 2 schemes introduced maximise the codecs' error resilience at the cost of slightly reduced video quality under error-free conditions. Gain-cost quantised, fixed but arbitrarily programmable rate discrete cosine transformed (DCT) video codecs, vector-quantised (VQ) and quad-tree (QT) coded algorithms are proposed and their video quality, complexity, compression ratio and error resilience trade-offs are comparatively analysed under identical conditions. Finally, our candidate codecs are compared to the standard H261, H.263 and MPEG2 benchmark codecs.

1. OVERVIEW OF VIDEO CODECS

In parallel to the rapid proliferation of multimedia communications services over fixed networks there is a growing demand for the added convenience of tetherless roaming. The earliest embodiment of the wireless multimedia communicator is expected to take the shape of a wireless videophone, which is likely to evolve from the family of the existing second generation mobile phones, such as the Pan-European GSM scheme [1], the North-American IS-54 [2] and IS-95 [3] communicators or the Japanese system [4]. Hence there is an urgent need for appropriate video codecs, which - *in contrast to existing standard codecs using bandwidth-efficient, but error-sensitive variable-length coding techniques*, such as the MPEG1 and MPEG2 arrangements [5, 6] or the ITU H261 and H263 codecs - can dispense with variable-length coding and hence produce an error resilient, very low rate stream that can be accommodated in the speech channel of the systems mentioned and can tolerate the bursty errors of the associated hostile wireless propagation environment [7]. Specifically, a bitrate of around 10 kbit/s is required by the speech channels of these systems and video users can be accommodated by an additional speech channel. Furthermore, the codecs have to lend themselves to fixed, but programmable rate operation in the intelligent multimode terminals (IMTs) of the near future, which will allow tele-traffic or channel-quality motivated transceiver reconfiguration [8, 9, 10, 11, 12].

The theory and practice of image compression has been consolidated in a number of established monographs by Netravali and Haskell [13], Jain [14], Jayant and Noll [15]. A plethora of video codecs have been proposed in the excellent special issues edited by Tzou, Mussmann and Aigawa [16], by Hubing [17] and Girod et al [18] for a range of bitrates and applications, but the individual contributions by a number of renowned authors are too numerous to review. Khansari, Jalali, Dubois and Mermelstein [19] as well as Mann Pelz [20] reported promising results on adopting the H261 codec for wireless applications by invoking powerful signal processing and error-control techniques in order to remedy the inherent source coding problems due to stretching its application domain to hostile wireless environments. Further important contributions in the field were due to Chen et al [21], Illgner and Lappe [22] Zhang [23], Ibaraki, Fujimoto and Nakano [24], Watanabe et al [25] etc, the MPEG4 consortium's endeavours [28], the efforts of the mobile audio-video terminal (MAVT) consortium. The European Advanced Communication Technologies and Services (ACTS) AC103 programme

[1] This treatise is complemented by a demonstration package portraying video sequences at various bit rates, which is down-loadable from http://www-mobile.ecs.soton.ac.uk or via http://www.ecs.soton.ac.uk.

has recently been launched with similar objectives.

The H.263 |27, 28] scheme is in many respects similar to the H.261 codec [29], but it incorporates a number of recent advances in the field, such as using half pixel resolution in the motion compensation, or configuring the codec for a lower datarate or better error resilience. Furthermore, four so-called negotiable coding options were introduced, which are referred to in H.263 parlance as Unrestricted Motion Vectors, Syntax-based arithmetic coding, Advance prediction, and predicted- (P) as well as bi-directional (B) frames in order to improve the coding performance. It also supports five different video resolutions. The bitstream generated by the H.261 and H.263 codecs is structured in a number of hierarchical layers, including the so-called picture layer, group of block layer, macroblock layer and block layer. In order to allow a high grade of flexibility to adopt to a various images, each of these layers have a self-descriptive structure commencing with a header segment in the upper three layers, allowing the codec to re-synchronise after loss of synchronisation following transmission errors. The received bitstream typically informs the decoder as to the inter- or intra-coded nature of a frame, the video resolution used, whether to expect any more information of a certain type, the index of the currently used quantiser, the location of encoded blocks containing large transform coefficients etc. This scheme achieves a high compression ratio over channels exhibiting a low bit error rate, but since the bit stream is self-descriptive, it is also rather vulnerable to channel errors. In [30] we reported on the design of a low-rate video transceiver, where the H.263 codec was constrained to operate at a constant bitrate using an appropriate packetisation algorithm, which adjusted the quantiser such that it would output the required number of bits per frame. Furthermore, using a low-rate feedback channel we "froze" the contents of the local and remote decoder, when transmission errors occured, which allowed the codec to operate over more hostile wireless channels.

As a design alternative, in this treatise we attempt to offer a comparative study of a range of fixed but arbitrarily programmable-rate 176 × 144 pixel head-and-shoulder Quarter Common Intermediate Format (QCIF) video codecs specially designed for videotelephony over existing and future mobile radio speech systems on the basis of a recent research programme [27, 31, 32, 33, 34, 35]. These codecs employ novel motion compensation and -tracking techniques as well as video "activity" identification and tracking. Various motion compensated error residual coding techniques are compared, which dispense with the self-descriptive, zig-zag scanning and run-length coding principle of the H.261 and H.263 codecs for the sake of higher activity tracking ability and improved robustness. For the same reasons, no motion vector estimation is used. Within the implementational complexity- and bitrate limits always an optimum constant bit allocation is sought.

The outline of the paper is as follows. Section 2 provides a comparative study of the appropriate motion compensation (MC) techniques, while section 3 proposes gain-cost quantised, fixed but arbitrarily programmable rate discrete cosine transformed (DCT) video codecs. Sections 4 and 5 offer a similar comparative study of vector-quantised (VQ) and quad-tree (QT) codecs. Our conclusions are presented in section 6.

2. Motion compensation

2.1. *Comparison of motion compensation techniques*

In video sequence coding, such as video telephony, typically a combination of temporal and spatial coding techniques is used in order to remove the "predictable or redundant" image contents and encode only the "unpredictable" information. Consecutive image frames f_n and f_{n-1} usually exhibit both intra- and inter-frame correlation. Most of the inter-frame redundancy can be removed by motion compensation (MC), where the motion compensated error residual (MCER), also often termed as displaced frame difference (DFD), is generated by subtracting the appropriately motion-translated blocks of the previous frame from the current one. How this can be achieved is the subject of this section.

The MCER typically contains a zero mean signal of non stationary nature. Fig. 1 depicts the two dimensional autocorrelation function (ACF) of a typical MCER and it's probability density function (PDF). The ACF of the MCER suggests that there is residual spatial correlation or predictability over a number adjacent pixels in both vertical and horizontal directions. This residual spatial correlation can be exploited using a variety of coding schemes. The PDF of the MCER signal in Fig. 1 also suggests that most of the values of the MCER are within a range of [-10, 10|, assuming that the source frames are sampled using an 8 bit resolution within the range of [0..255|. However, when using a lossy MCER coding scheme, the quantisation distortion associated with the MCER encoder will accumulate in the local reconstructed frame buffer and it will alter the statistical characteristics of the forthcoming MCER frames, reducing the correlation and resulting in a more widely spread PDF.

Recently a variety of MC algorithms have been published in the literature [36, 37, 38, 39]. Many of them constitute variants of the well-known full search algorithm. The so-called grid interpolation techniques introduced quite recently by Huang et al and Moisala et al are rooted in model based techniques [40, 41], while Girod improved the accuracy of the motion prediction by interpolating the given images to a higher resolution and defining the motion vectors (MVs) at a sub-pixel accuracy |42|. The family of the so-called pel-recursive techniques addressed by Huang et al and Efstratiadis et al |43, 44] predict the motion vectors for the current frame by observing the speed and orientation of each pel in previous frames. Thus, no additional information must be transmitted for

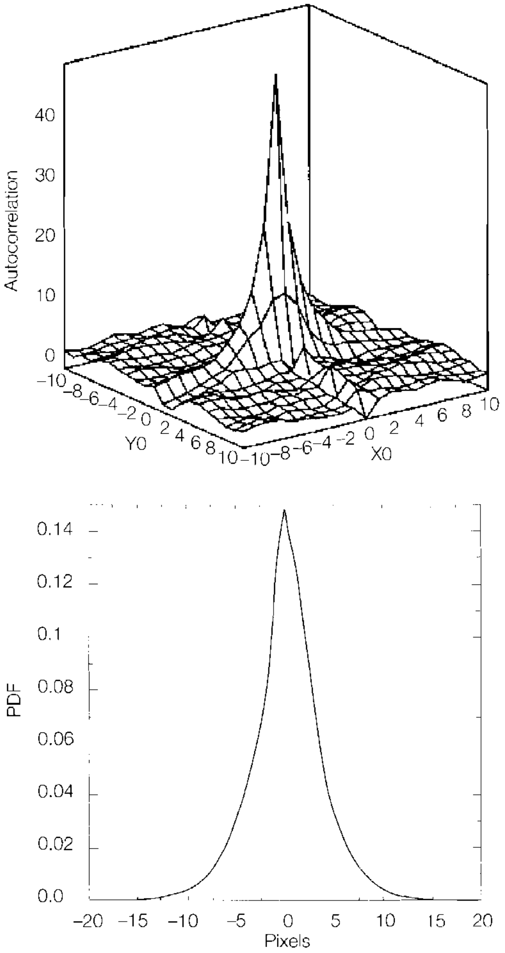


Fig. 1 - Two dimensional autocorrelation function and the PDF of the MCER for QCIF videophone sequences.

the motion vectors, because the prediction is based on frames which are available at both the transmitter and the receiver at the same time. However, due to its error sensitivity this technique is not suited for wireless videophony. Since the choice of the MC search algorithm has a profound effect on both the associated search complexity and the video quality achieved, in our forthcoming deliberations we compare the performance of a range of MC techniques. We refrained however from using the sub-pixel resolution and error-sensitive predictive motion vector coding techniques of the H.263 codec, where the motion vectors can be generated either on a block-by-block or on a macroblock basis, which refers to four blocks.

The full search MC algorithm evaluates the associated matching criterion, in our case the mean squared error (mse) for every possible motion vector (MV) within the given search scope and selects that particular one, which results in the lowest matching error. We note however that in many applications the so-called mean absolut error matching criterion is preferred due to its evaluation simplicity.

The so-called *tree or hierarchical search algorithm* proposed by Jain and Jain and adopted by Wang et al [45, 46] drastically reduces the complexity of MC while maintaining a high quality motion prediction. As demonstrated by Fig. 2, the associated complexity reduction in comparison to the full-search scheme is achieved without a significant loss in terms of PSNR performance. The set of curves displayed in Fig. 2 demonstrates the performance of various search algorithms applied to a typical head-and-shoulder video sequence. A total of 50 active motion vectors were assumed out of the 396 8 × 8 QCIF blocks, while for the remaining 346 blocks simple frame differencing was used. The associated MCER was not transmitted to the decoder at all, which was equivalent to the assumption of MCER = 0. This crude assumption represented a worst-case codec performance, which was found adequate to assess the performance of the various MC schemes.

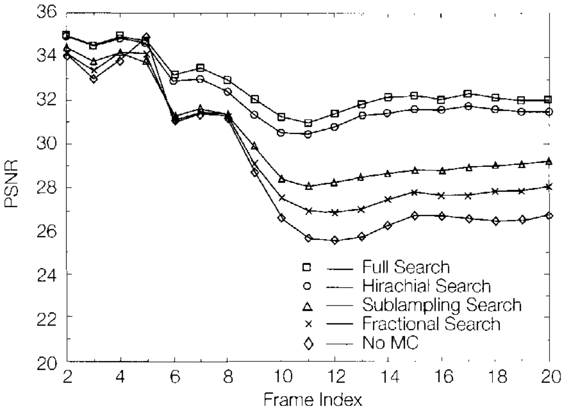


Fig. 2 - Performance comparison of various QCIF MC algorithms in terms of PSNR versus frameindex when no encoding of the MCER is performed for the "Miss America" sequence.

The *subsampling search* algorithm operates similarly to the above tree search. However, instead of reducing the search window in consecutive stages, the whole search area is subsampled by a given factor. The search in the lower resolution domain is less complex, since the number of possible locations and the block size are reduced according to the subsampling factor. Once the best location has been found in the lower resolution domain, the resolution will be increased step by step. The advantage is, that the search window can be limited around the best position found in the previous step. The procedure continues until the original resolution is reached. Although the search process is similar to the hierarchical search, it does not reach its performance, as demonstrated by Fig. 2. Let us now concentrate on how efficient gain-cost motivated MC can be achieved in fixed-rate video codecs, where the expression "gain-cost motivated" implies that the MV encoding bits representing the coding "cost" are allocated on the basis of striving for maximum coding "gain".

2.2. *Gain-cost controlled motion compensation*

Motion compensation reduces the MCER energy at the cost of additional complexity and channel capacity demand for the motion vectors. The contribution of each vector towards the MCER entropy reduction strongly depends on the movement which is associated with each vector. For most of the blocks the MCER

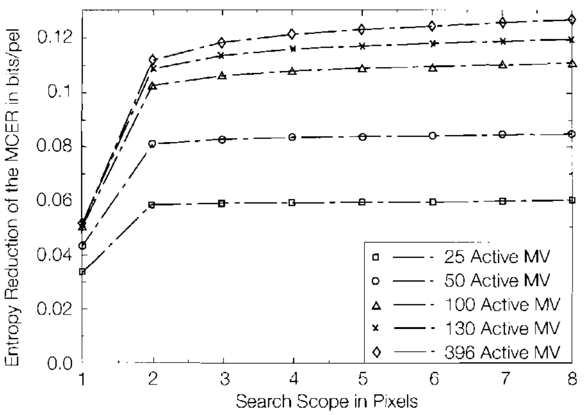


Fig. 3 - Entropy reduction of the MCER as a function of the search scope and the number of active MVs for QCIF head-and-shoulders sequences (Miss America).

entropy reduction does not justify the additional transmission overhead associated with the MV. Fig. 3 reveals this entropy reduction for various numbers of active MVs based on QCIF frames containing a total of 396 8 × 8 blocks. In order to evaluate the potential gain of every motion vector, we measured the MCER energy reduction and selected the most efficient vectors. All passive MVs were set to [0, 0] and simple frame differencing is used at "passive" locations. Surprisingly, our results suggested that the most important 25% of the MVs, ie using about 100 active MVs, resulted in 80 – 90% of the possible entropy reduction.

The disadvantage of the above motion passive / active concept is the additional requirement of indeces or tables necessary to identify the active vectors. In a conceptually simple approach we can assign a one-bit motion activity flag to each of the 396 8 × 8 blocks, which constitutes a transmission overhead of 396 bits for each QCIF frame. A very efficient compaction technique will be introduced in section 3.7 in order to further compress this motion activity table and hence reduce the average number of bits per MV. Fig. 4 reveals the required coding rate of the MVs, when

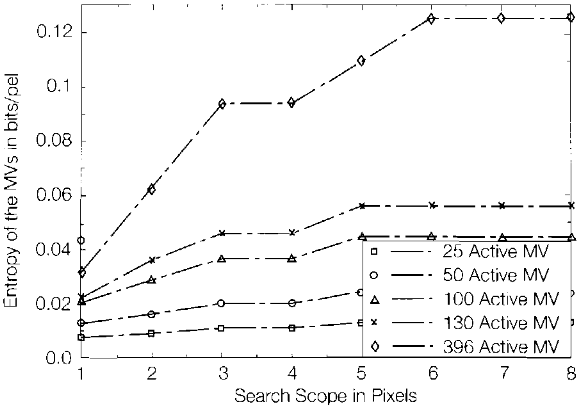


Fig. 4 - Required coding rate for the MVs as function of search scope and number of active MVs in case of QCIF head-and-shoulders sequences (Miss America).

applying the above mentioned efficient activity table compression method. From Figs. 3 and 4 we concluded that the optimum search window size in terms of overall entropy reduction is between 2 × 2 and 3 × 3 pels, while the adequate number of active motion vectors is below 100 in case of QCIF images. When using more active MVs, the number of bits assigned to their encoding will reduce the number of bits available for the encoding of the MCER and inevitably reduce the image quality in case of the fixed bit rate budget of the targeted mobile radio speech channels.

Having compared a range of block matching motion estimation techniques, we found that in case of QCIF images upon using a rather limited proportion of active motion vectors we were able to reduce the DFD energy significantly. Another result was that small search window sizes of 4 × 4 pixels were adequate to maintain a low complexity and near-optimum performance. Such a small search window allowed us to use the highest complexity exhaustive search method, while keeping the technique's computational demand within the requirements of real time applications. This contributes towards practical VLSI implementations. Finally we note that backward estimation techniques, such as pel-recursive motion compensation, or the entropy-coded variable-length predictive motion vector estimation technique of the H.263 codec are not suitable for communications over high error rate channels.

In accordance with our initial goal, a salient feature of the proposed cost/gain controlled MC scheme is that it lends itself to tele-traffic and/or channel quality motivated adaptive reconfiguration, which is an essential element of the so-called 3-rd generation wireless intelligent multimode terminal (IMT) of the near future. Although research in this field is rapidly progressing [8 - 11, 12], the existing 2-nd generation systems, such as the Pan-European GSM system [1], as well as American IS-54, IS-95 and the Japanese systems can benefit from the proposed codecs by fixing the video bit rate generated to their speech rate and provide video telephone services. Similar programmable-rate coding techniques will be sought during our further discourse for the encoding of the MCER.

The MCER can be encoded using a range of techniques, including subband coding (SBC), which was treated by Woods and O'Neil [47, 48], Gharavi and Tabatabai [49, 50, 51], Ngan [52], Podilchuk, Jayant and Farvardin [53], wavelet coding documented for example by Katto et al [54], Discrete Cosine Transformation [5, 6, 14, 34] (DCT), vector quantisation (VQ) [55, 56] or Quad-tree [57, 58, 59] (QT) coding. Most of these schemes share the structure of Fig. 7, which will be described in depth in the forthcoming sections. Observe at the top section of Fig. 7 that an optional model-based parametric coding (PC) scheme can also be invoked in the subjectively most important eye and lip regions of the image. We continue our discourse by proposing a suite of DCT-based codecs.

3. Low bitrate DCT codecs

3.1. *Background*

A range of orthogonal transforms have been suggested in the literature for various applications [14, 60]. The discrete cosine transform (DCT) is widely used for image compression, since its compression ratio approaches that of the optimum Karhunen-Loeve (KL) transform [14], while maintaining a significantly lower implementational complexity. This is due to the fact that the KL transform's so-called basis vectors used in the transformation are data dependent and hence they have to be computed for each data segment, before transformation can take place. The simple reason for the so-called energy compaction property of orthogonal transforms, in particular that of the DCT is that for correlated time-domain sequences, which have a flat-centred autocorrelation function, the spatial frequency-domain representative tends to be of low-pass nature, where most of the image energy is conveyed by a small proportion of the coefficients.

The efficiency of the DCT transform has constantly attracted the interest of researchers and recently many attractive schemes have been published by Mirchandani and Palau [61], Yamaguchi [62], Ngan [63], Clarke [64], Wang et al [65], Azari and Ghanbari [66], etc. These approaches are typically associated with a fluctuating, time-variant compression ratio, requiring adaptive feedback controlled buffering methods in order to allow a constant bit rate. The following MCER coding approach presents a fixed- but arbitrarily programmable rate DCT scheme, which can dispense with adaptive feedback buffering and lends itself to applications, such as the wireless IMT.

3.2. *The quantisers*

In order to adequately represent the MCER, appropriately trained minimum mse Max-Lloyd quantisers [14] are required. Once a spatially correlated MCER block $b(i, j)$ of, for example, 8 × 8 pixels is orthogonally transformed using the DCT matrix T defined in eq. (1) as:

$$T_{mn} = \frac{2c(m)c(n)}{N} \cdot \sum_{i=0}^{b}\sum_{j=0}^{b} b(i,j)\cos\frac{(2i+1)m\pi}{2b}\cos\frac{(2j+1)n\pi}{2b} \qquad (1)$$

$$c(m) = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } (m=0) \\ 1 & \text{otherwise} \end{cases}$$

in the spatial frequency domain, most of its energy is typically concentrated in the low horizontal and vertical frequency regions corresponding to the "top left hand corner region" of its two-dimensional spectrum, if the source signal subjected to DCT exhibits a high correlation. In case of using motion compensation much of this correlation or predictability is removed and hence often the MCER signal has an expanded spatial frequency domain spread. The H.261 and H.263 codecs hence invoke a zig-zag scanning technique combined with appropriate run-length coding tables, in order to cater for various frequency-domain energy distributions. Although in bandwidth-efficiency terms this is an attractive technique, due to the paucity of bits in very low-rate codecs, where typically only a small fraction of the high-energy coefficients can realistically be encoded, a zonal mask often gives a similar quality at a concomitant higher robustness.

Based on these arguments we opted for the latter zonal mask, assigning higher resolution quantisers in the high-energy frequency regions, whereas lower energy frequency regions are quantised with more coarse quantisers or completely neglected by setting them to zero. We note again, however that due to the MCER's non-stationary statistical properties often an uncorrelated MCER is resulted, for which a more widely spread frequency domain representation is yielded. As mentioned, in these cases the DCT may in fact result in energy expansion, rather than compaction, which will be catered for by using an appropriate multi-class quantiser design. Clearly, the issue of choosing the appropriate frequency regions for quantisation and the appropriate quantiser resolution are the most crucial design steps for a DCT codec. The next two Sections report on how we derived an initial set of quantisers, and how we expanded this initial experience to design a set of final quantisers for a quad-class scheme.

3.3. *Quantiser Training for Single-Class DCT*

Initially we embarked on determining, which of the DCT coefficients had to be retained and quantised, and which could be ignored or set to zero, while not inflicting severe image quality degradation. We commenced by setting some of the DCT coefficients to zero, whereas the remaining coefficients were left unquantised. We found, that in order to maintain adequate videophone quality, about six coefficients were needed, which were concentrated in the low horizontal and vertical frequency regions.

In a second step, we assigned the appropriate number of reconstruction levels, $n_{\text{rec}} = 2^k \,|\, k = \{1,2,3,\ldots\}$, so that each coefficient was quantised to k bits, while maintaining the required target bit rate. Initially we allocated a generous eight bits for the DC coefficient, four bits for each of the positions [0,1] and [1,0] on the [8,8]-dimensional frequency plane, and two bits for each of the positions [2,1],[1,1] and [1,2]. Appropriately trained Max-Lloyd quantiser were used to achieve minimum-distortion quantisation. Similarly to the gain / cost controlled motion active / passive classification, we obtained the appropriate training data for the DCT Max-Lloyd quantisers (MLQ) by recording the DCT coefficients of the MCER for the 30 highest-energy blocks of each 396-block frame for a 100 frames long motion compensated sequence. This selection of training blocks considers explicitly the fact that, similarly to MC, we decided to incorporate the active / passive classification scheme also into the MCER coding scheme, which assists us in adaptively controlling the bit rate by adjusting the proportion of motion- and DCT-active blocks. Perceptually pleasant videophone quality was achieved, when quantising every 8 × 8 pixel block to a total of 22 bits. In order to achieve the maximum possible compression ratio attainable using this technique, while maintaining an adequate image quality, we iteratively reduced the number of reconstruction levels until we reached the required image quality. At this stage the entire block was quantised to a total of ten bits, which led to the bit allocation depicted in Table 1.

Table 1 - *Bit allocation table for the single-class quantiser*

3	2	1	0	0	0	0	0
2	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

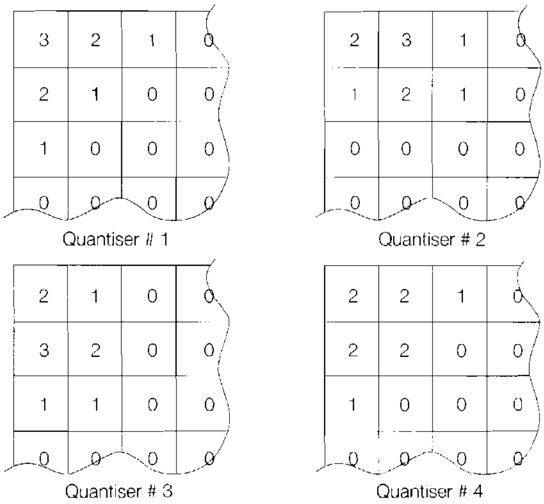
3.4. *Quantiser training for multi-class DCT*

When using the above single-class bit allocation, in case of uncorrelated MCER sequences noticeable degradation was caused at the edges of motion trajectories or at areas of high contrast. Furthermore, if there is more correlation in one particular spatial direction than the other, the centre of the spectral-domain energy peak moves towards the direction of higher correlation in time domain. This happens if the block contains for example a striped texture. Hence we added two sets of quantisers, which shall cater for blocks with higher horizontal or vertical correlation. A fourth quantiser was added for the encoding of less correlated time-domain, ie more spread frequency-domain sequences, having a more uniformly distributed bit allocation. The bit allocation for all four quantisers is depicted in Fig. 5. Again, we emphasize that we dispensed with vulnerable zig-zag scanning and variable-length entropy-based DCT coefficient coding, which achieves in this low-rate application only marginal gains at the cost of higher error sensitivity.

After fixing the quantiser resolutions we had to find an appropriate training sequence for each set of quantisers. Hence a classification algorithm was needed in order to assign each of the training blocks in to one of those categories. We decided to use a quantisation function, which we will refer to as a "pseudo-quantiser", producing a randomly quantised quantity $\tilde{q}$ from the unquantised variable q by contaminating q using an additive, quantiser-resolution dependent fraction of the unquantised variable itself. Specifically, we attempted to simulate each quantiser by applying eq. (2),

$$\tilde{q} = q \times \left[1.0 + 1.5 \frac{(-1)^{\mathrm{RAND}}}{n} \right] \qquad (2)$$

where q is the quantity to be quantised, $\tilde{q}$ represents its "pseudo-quantised" value, n denotes the number of simulated reconstruction levels and *RAND* is a random integer. Clearly, this "pseudo-quantiser" simulates the quantisation error by randomly adding or subtracting a fraction of itself. The magnitude of the added distortion is inversely proportional to the number of quantisation levels n.



3	2	1	0
2	1	0	0
1	0	0	0
0	0	0	0

Quantiser # 1

2	3	1	0
1	2	1	0
0	0	0	0
0	0	0	0

Quantiser # 2

2	1	0	0
3	2	0	0
1	1	0	0
0	0	0	0

Quantiser # 3

2	2	1	0
2	2	0	0
1	0	0	0
0	0	0	0

Quantiser # 4

Fig. 5 - Bit allocation for the Quad-Class QCIF DCT quantisers.

Each quantiser training block was DCT transformed, pseudo quantised and then inverse transformed back to the time domain, while using all of the four sets of quantiser resolutions seen in Fig. 5. The unquantised coefficients corresponding to the specific quantiser of Fig. 5 leading to the best reconstruction in the minimum mse (mmse) sense were collected separately for each class. Finally, the Max-Lloyd quantisers for each energy distribution class were derived from each of these training sets. Surprisingly, the initial set of quantisers performed so well that retraining the quantisers did not result in any significant improvement.

Fig. 6 exemplifies the codec's PSNR performance for a single-, dual- and quad-class DCT codec at a constant bit rate of 800 bits per frame using the "Claire" sequence, which will be used throughout this treatise for the sake of comparability of the various schemes studied, unless otherwise stated. At this stage it is convenient to fix the adaptive IMT codec's bit rate to a constant value in the centre of its expected operating range. This rate is slightly more than the 6.7 and 8 kbit/s Vector Sum Excited Linear Predictive [67] (VSELP) speech rate of

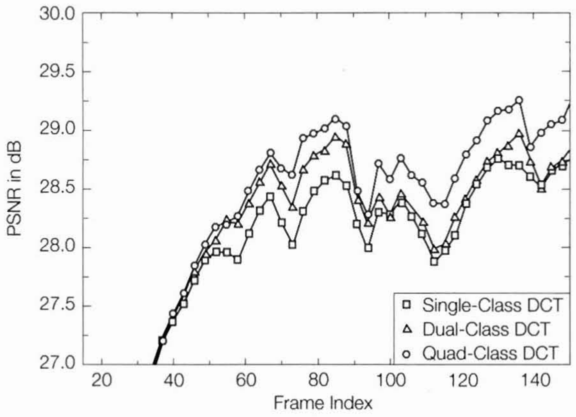


Fig. 6 - PSNR versus frame index performance of various DCT quantiser schemes at 8 kbit/s for the "Claire" sequence.

the Japanese PDC [4] and the American IS-54 systems [2], respectively, about the same as the 9.6 kbit/s active voice rate of the IS-95 system [3] and slightly less than the Pan-European GSM [1] system's 13 kbit/s Regular Pulse Excited [67] (RPE) voice rate.

The twin-class codec was using the quantiser sets derived for high and normal energy compaction which are depicted in the top left and bottom right corner of Fig. 5. Note that the twin-class scheme requires one, and the quad-class scheme two more bits per DCT transformed block for the encoding of the energy distribution classifier. Hence the number of active blocks to be encoded had to be reduced, when the overall bit rate was kept constant, nonetheless, the preceived image quality increased significantly and the objective PSNR also improved by up to 0.7 dB due to the classification technique. The relative frequency of selecting any one of the four quantisers varies from frame to frame and also depends on the specific encoded sequences. The average relative frequency of the four quantisers was evaluated using the "Miss America" sequence for both the twin- and the quad-class DCT schemes and it is shown in Table 2. Observe that the highest relative frequency bottom entry of the quad-class DCT codec in Table 2 is associated with an extended horizontal correlation, which is also confirmed by the autocorrelation function of the typical MCER, as seen in Fig. 1.

Table 2 - *Relative frequency of the various quantisers in multi-class DCT codecs*

Quantiser Set Type	Relative Frequ. Quad-Cl. DCT	Relative Frequ. Dual-Cl. DCT
Very High Correl.	14.7%	39%
High Correl.	29.3%	61%
Increased Vert. C.	24.0%	-
Increased Hor. C.	32.0%	-

3.5. *DCT codec outline*

The proposed programmable codec was designed to switch between intra- and inter-frame modes of operation, as seen in Fig. 7. This was necessary because the DCT based inter frame codec assumes that the previous reconstructed frame is known to both the encoder and decoder. At the commencement of communications the reconstructed frame buffers of the encoder and decoder are empty, hence the MCER energy for the first frame would be very high, constituted by the first frame to be encoded. The quantisers we derived in section 3.2 are not trained for this particular case, and hence the DC-coefficient quantiser often would be driven to saturation. Therefore we transmit a single independent intra-frame coded frame during the initial phase of communication in order to replenish the reconstructed frame buffers. Once switched to inter-frame mode, any further mode switches are optional and only required if a drastic change of scene occurs.

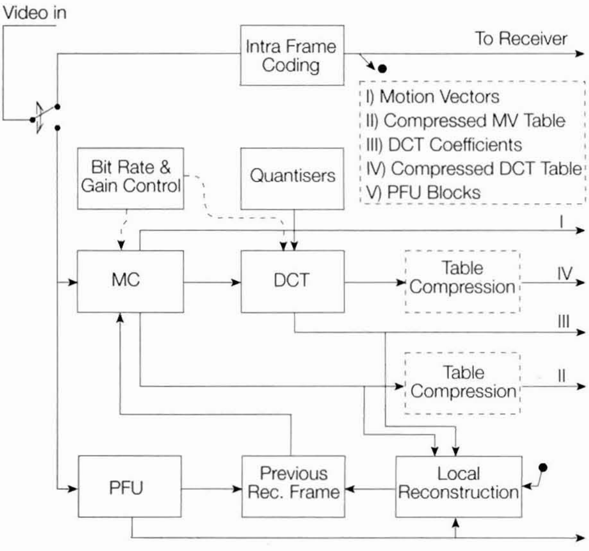


Fig. 7 - Schematic of the multi-class DCT codec.

In the inter-frame mode the incoming video sequence is first motion compensated and then the resulting MCER is DCT encoded. The active / passive decisions carried out during both the MC and DCT coding stages are cost gain controlled within the allocated constant, but programmable bit rate budget. The optional table compression algorithm printed in broken lines in Fig. 7 will be described at a later stage, and it is incorporated in order to reduce the amount of redundancy associated with the transmission of the 396-entry motion- and DCT-activity tables. Finally, a partial forced update (PFU) algorithm is included in order to ensure that a realignment of the encoder's and decoder's reconstructed buffers becomes possible, while operating under high error-rate channel conditions. In the forthcoming Sections we will detail the role of each block in Fig. 7.

3.6. *Initial intra frame coding*

The objective of the intra-frame mode is to provide an initial frame for both the encoder's and decoder's reconstructed frame buffer in order to ensure that the initial MCER frame does not saturate the DC-coefficient quantiser. This can be achieved by dividing the frame into perfectly tiling blocks and coarsely quantising the average of each block. It was found appropriate to quantise each block average to sixteen levels, which are equi-distantly distributed between the absolute pixel values of 52 and 216. Since the entire intra-frame coded image must be encoded within the limits of the per frame bit rate budget, the block size had to be adjusted accordingly. The resulting block sizes for various bit rates are depicted in Table 3. When the intra-frame coded block size did not tile the QCIF format perfectly, the codec increased the block size for those particular blocks which were situated along the fringes of the frame.

Table 3 - *Blocks size versus bitnumber/frame for the QCIF intra-frame codec*

Bits per Frame	Block Size
500	14 × 14
800	12 × 12
1000	10 × 10
1200	10 × 10
1500	9 × 9

3.7. *Active / passive table compression*

On the basis of section 2 the adaptive bitrate control algorithm selects the specific number of most efficient MVs according to the available bitrate budget. The remaining MVs are set to zero, before subtracting the current motion translated reconstructed block from the incoming one, which corresponds to simple frame differencing. As mentioned, the conceptually simplest solution to convey the motion- and DCT-activity information would be to transmit a nine-bit address along with every active block of the 396-block QCIF frame. Assuming for example a total of 50 + 50 = 100 motion- and DCT active blocks, 900 bits/frame would be necessary only to identify their locations, which would nearly exhaust the targeted bitrate budget without encoding the MCER. The channel capacity requirement of this scheme is represented by the "No Tables" curve in Fig. 8.

As alluded to before, it is a more convenient approach to establish an active / passive table, containing for each of the 396 blocks a one bit flag marking the corresponding block either as "active" or "passive". In this case the total number of bits conveying the active / passive block information for a whole frame is 396, irrespective of the actual motion- or DCT-activity. The bitrate requirement of this scenario is characterised by the "Fixed Table" legend in Fig. 8.

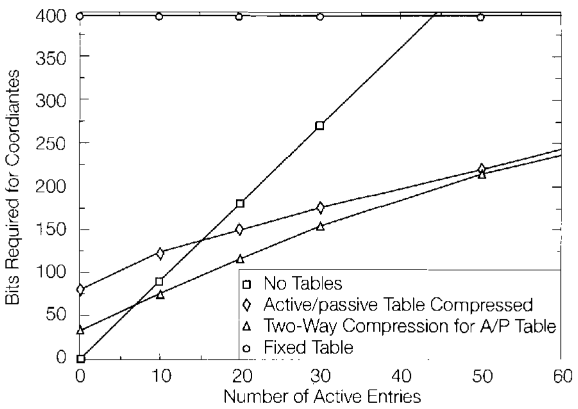


Fig. 8 - Comparison of various active / passive classifier encoding schemes.

The coding efficiency of the above scenario can be further improved bearing in mind that the relative frequency of active blocks for very low bit rate applications is usually below 15%, in which case most of the entries of the active / passive table will be marked as "zero". This implies predictability or redundancy, which can be removed by run length or entropy encoding. Hence we decided to scan a certain number of the activity flag bits from left-to-right and top-to-bottom into symbols, which we then Huffman encoded. We found that the optimum number of bits per symbol was around five. This bit to symbol conversion exploits most of the latent redundancy inherent in the bit-by-bit representation, while preventing us from optimising the code words of the symbol-based Huffman codec for a certain activity rate. The "Active/Passive Table compressed" curve of Fig. 8 shows the required bitrate of this technique.

However, this technique does not exploit the fact that increased motion activity often covers a group of several blocks rather than a single block. Therefore we further developed the above approach, such that the motion activity flags were grouped into two-by-two blocks, covering four original 8 × 8 pixel blocks. For those groups that did contain active motion flags, we assigned a four-bit activity symbol, reflecting its activity-flag contents. These symbols were then Huffman coded. This concept of only transmitting information regarding those groups of vectors which were active required a second active / passive (*A/P*) table. This second table reflected, which of the grouped motion activity flags were active. It contained 396/4 = 99 entries, which we again packetised to three entries per symbol, a value that was found to guarantee best coding efficiency, and Huffman encoded. This method was found superior to the previous technique, in particular, when the proportion of active blocks was constrained to below 10% or 40 blocks per frame, as demonstrated by the "Two-way compression for *A/P* Table" curve of Fig. 8. Observe in the Figure that for example at 50 active blocks less than

250 bits is required to convey the *A*/*P* block information for the whole frame, corresponding to a total of less than 500 bits for motion- and DCT-activity.

3.8. *Partial forced update of the reconstructed frame buffers*

Since inter-frame codecs are prone to error propagation due to the misalignment between the reconstructed frame buffer of the encoder and decoder we invoked a combination of "leakage- and forced update techniques". Specifically, we selected a predetermined, bit rate dependent number of 8 × 8 pixel blocks in the incoming frame buffer, for which we determined the coarsely quantised block-average as it was detailed in section 3.6. Then every pixel of the selected blocks in both the encoder's and decoder's reconstructed frame buffers was first scaled down by a certain leakage factor l and the 4 bit quantised block average was superimposed, which was scaled down by the factor $(1 - l)$. This partial forced updating (PFU) scheme guarantees that the average MCER energy does not change and the DC-quantisers of the subsequent DCT scheme are not overloaded. The leakage factor values used in our approach are typically between 0.3 and 0.5 in order to ensure that the PFU blocks do not become blurred in the next frame.

The PSNR degradation inflicted by this PFU technique amongst error free conditions is portrayed in Fig. 9, which is modest, as long as the proportion of PFU blocks is low. For example, when a PFU rate of about 10% corresponding to 40 blocks is applied, the average PSNR of the "Claire" sequence at a coding rate of 9.6 kbit/s degrades by about 0.5 dB at a concomitantly improved codec robustness. However, as the update rate is increased to around 13%, the PSNR degradation becomes more noticable. We used the same block-average quantiser, which was invoked for intra frame coding in section 3.6, but the size of the updated blocks was fixed to 8 × 8 pixels. The number of updated blocks per frame and the leakage factor determine the inflicted intentional video impairment due to PFU and the codec's ability to recover from erroneous conditions, ie the codec's robustness. Depending on the bitrate, we typically limited the number of updated blocks per frame to below 20 out of 396 or about 5%, due to the low associated PSNR reduction and bitrate contribution.

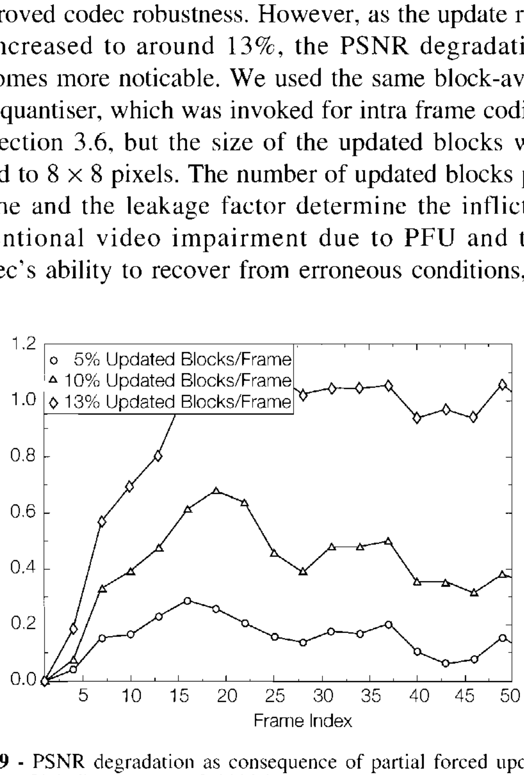


Fig. 9 - PSNR degradation as consequence of partial forced update for the "Claire" sequence at 9.6 kbit/s.

3.9. *The gain/cost controlled inter frame codec*

In the previous sections we described the components of the coding scheme. This section reports on how we combined the introduced components in order to achieve a high frame rate coding scheme for QCIF images at bit rates at and below 10 kbit/s.

The motion detector attempts to minimise the MCER between the incoming and the local reconstructed frame buffers. As a result of the conclusions of section 2.2 the search window is limited to a size of 4 × 4 pixels, which limits the search scope to –2 in negative and +1 in the positive horizontal / vertical spatial directions. The motion detector stores the best motion vector for each of the 396 blocks as well as the corresponding motion compensation gain. Now the bit rate control unit marks inefficient motion vectors as passive and determines the corresponding active / passive tables as described in section 3.7 and computes the resulting overall bit rate requirement for the MVs. The codec then estimates how many encoded MVs can be accommodated by the available bit rate budget and relegates some of the active motion vectors to the motion passive class. Then the bit rate requirement is updated and compared to the bit rate limit. This de-activation of motion vectors continues until we reach a predetermined maximum bit rate for the MC.

Finally, motion compensation takes place for the motion active blocks, while for the passive blocks simple frame differencing is applied. The MCER is passed on to the DCT codec, which has a similar structure to the motion compensation scheme. Each block is transformed to the frequency domain, quantised and then transformed back to the time domain in order to assess the potential benefit of the DCT coding in terms of overall PSNR or mse contribution. The best set of quantisers is found by a full search, ie by invoking all the four available quantisers and evaluating their mse performance. Now, the bit rate control unit determines the number of available bits for the DCT codec, remaining from the total budget for the frame after reserving the required capacity for the partial forced update, the motion compensation and the so-called frame alignment word (FAW). Again the active / passive DCT tables are determined and blocks attaining low mse gains are not encoded in order to meet the overall bit rate requirement. In order to emphasize the subjectively more important central eye and lip region of the screen the codec incorporates the option of scaling the DCT gains and allows for the codec to gradually improve its image representation in its centeral section. This is particularly important, when operating at 5 – 8 kbit/s or during the first transmitted frames, while the codec builds up fine details in order to reach its steady-state video quality, commencing from the coarsely quantised intra-frame coded initial state.

3.10. *Complexity reduction techniques*

The number of mutiplications required for the DCT of a $b \times b$ image block or matrix B is proportional to b^4 [14]. The dimensionality of the problem can be reduced to a more realistic b^3, if we invoke the so-called separable transform: $B_{\mathrm{dct}} = TBT^t$, where T is the $b \times b$ unitary transformation matrix we defined by eq. (1) and T^t is the transpose of T. A single matrix multiplication of a $b \times b$ matrix requires b^2 $(2b - 1)$ Flops. In case of our 8×8 blocks ($b = 8$), 1920 Flops are necessary to evaluate one transformation. If we apply gain controlled quad-class DCT, it is necessary to transform each block back to the time domain after applying each of the four quantisers. If we apply the *mse* criterion to select the best of the possible four quantisers, we require a total of $10b^2$ $(2b - 1) + 3b^2 - 1 \simeq 20b^2$ Flops per block, which equals to 3.9 Mflop per QCIF frame. This can be reduced by about 25% upon exploiting some regularities in the matrix operations at the cost of increased data handling steps [68]. This may be important for a direct implementation in silicon.

A much more efficient way of reducing the computational complexity becomes possible using a block classification algorithm prior to any encoding steps. Similar to our experience in section 2.2 as regards to MC, we found that a substantial proportion of the MCER contains flat blocks which are unlikely to be selected by a gain-cost controlled codec. Hence, prior to the actual encoding the encoder determines the energy contents of all blocks. Only the blocks exhibiting high energy contents, i.e. blocks carrying important information, are considered for the MC and DCT encoding operations. In case of the MC, the encoder first determines the frame difference and then the motion prediction is focused on the locations where significant movement took place. After carrying out the MC step, a similar pre-classification phase is invoked for the DCT. The degradation caused by the above sub-optimum pre-selection phase is depicted in Fig. 10 for a range of active block proportions between 20 and 50%. These results reveal that the complexity can be reduced by about 50% without noticeable PSNR impairment, when compared to the original scheme. When the proportion of active blocks is as low as 20%, the pre-classification is more likely to be deceived than in case of retaining a higher proportion of blocks. When assuming for example a 38% block activity rate, the codec exhibits a complexity of 2 Mflop per frame or 20 Mflops, which is only marginally higher than that of the standard so-called half-rate GSM speech codec.

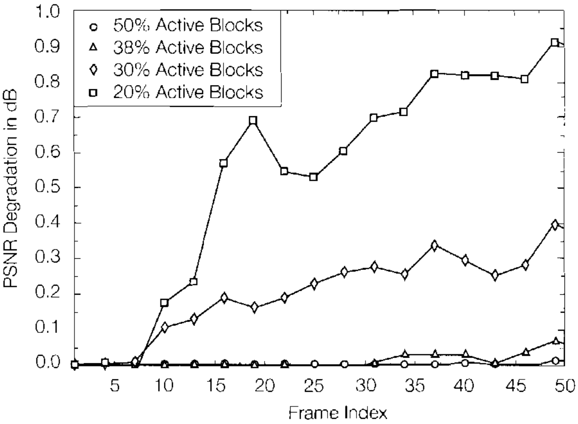


Fig. 10 - PSNR degradation caused by the preselection for the "Claire" sequence at 9.6 kbit/s.

3.11. *Adaptive bit allocation strategy*

The bit stream transmitted to the decoder consists of bits for:

- the averages for the partial forced update,
- the table for the active MVs,
- the active MVs,
- the table for the active DCT blocks and
- the active DCT coefficients,

all of which must be allocated adaptively. We found that the best subjective and objective videophone quality was achieved, when the number of active blocks for the MC and DCT was roughly the same, although not necessarily the same blocks were processed by the two independent algorithms. The encoding of the motion vectors requires only 4 bits per active block, while that of the DCT coefficients needs 12 bits/block, including the quantiser classifier. Hence we earmarked between 1/2 and 2/3 of the available bit rate budget to the DCT activity table and DCT coefficients, while the remaining bits were used for the MC and PFU. The PFU was typically configured to refresh 22 out of the 396 blocks in each frame. Therefore 4 × 22 = 88 bits were reserved for the PFU. The actual number of encoded DCT blocks and MVs depended on the selected bit rate and typically varied between 30 and 50 for bit rates between 8 and 12 kbit/s at a scanning rate of 10 frames/s.

The output of the codec contains two classes of bits. Namely, the entropy encoded MC- and DCT-activity tables on one hand, which constitute the more vulnerable Class 1, and the less sensitive Class 2 MV, DCT and PFU bits on the other hand. The first class of information is, due to the reliance of the encoding procedure on Huffman coding, extremely vulnerable against any corruption. A corrupted bit is likely to create a code associated with a different length and, as a result, the entire frame may have to be dropped or re-transmitted. In our further discourse we will refer to this DCT codec as DCTC1.

However, since the high vulnerability of the Huffman-coded DCTC1 to channel errors is unacceptable in some applications, we also contrived another, more robust codec, which sacrifices coding efficiency and abandons the Huffman coding concept for the sake of improved

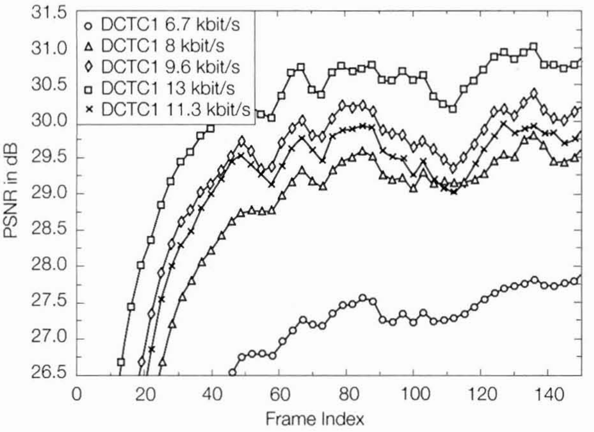


Fig. 11 - PSNR versus frame index performance of DCTC1 at various bit rates and for DCTC2 at 11.3 kbit/s for the "Claire" sequence.

error resilience. Explicitly, in DCTC2 we decided to transmit the index of each active DCT block and MV requiring 9 bits to identify one of the 396 indices using the so-called enumerative method. The increased robustness of the codec is associated with an approximately 35% increased bit rate. As Fig. 11 reveals, DCTC1 at 8 kbit/s achieves a similar quality to that of DCTC2 at 11.3 kbit/s.

3.12. *Performance of DCTC1 and DCTC2*

The performance of DCTC1 was tested at 10 frames/s and 6.7, 8, 9.6 and 13 kbit/s, which are the speech rates of the Japanese PDC [4], American IS-54 [2], IS-95 [3] and the Pan-European GSM system [1], respectively. The results for DCTC2 are similar at a 35% higher bit rate. The proposed DCTC1 achieved a similar PSNR to the CCITT H261 scheme investigated in [20] and [34] at 5 frames/s, while operating at twice the frame rate, namely at 10 frames/s, and maintaining a similar bit rate, which implies a factor two higher compression ratio. A more rigorous comparison will be offered in the Conclusions Section. The improved performance is mainly attributable to the intelligent cost-gain quantisation invoked throughout the coding operation, while maintaining a moderate complexity of about 20 Mflops [2].

3.13. *DCT codec robustness*

As regards to the codec's error sensitivity, we have to differentiate between two possible error events. If the run-length encoded Class 1 bits are corrupted, it is likely that a codeword of a different length is generated and the decoding process becomes corrupted. This error is often detectable, since the erroneously decoded frame length becomes different from the currently expected number of bits per video frame. Hence, a single bit error can force the decoder to drop an entire frame. If, however, one of the Class 2 PFU, DCT or MV bits is corrupted, the decoder is unable to detect the error event, but only a maximum of two blocks are affected by such a single bit error. The error sensitivity difference between the run-length and non-run-length encoded bits is highlighted in Fig. 12. If the whole bit stream of DCTC1 is subjected to random bit errors, a *BER* of $2 \cdot 10^{-4}$ is sufficient to inflict unacceptable video degradation. If, however, bit errors only affect the non-run-length encoded Class 2 bits, while the RL-coded bits remain intact, the codec can tolerate *BER*s up to $2 \cdot 10^{-2}$. In [31] we proposed an appropriate transmission scheme, which takes advantage of this characteristic. As evidenced by Fig. 12, the absence of run-length encoded bits increases the error resilience of DCTC2 by an order of magnitude. Therefore DCTC2 is better suited for example for mobile applications over Rayleigh fading channels. Further issues of un-equal protection FEC and ARQ schemes are discussed in [31].

Having studied the algorithmic and performance issues of DCT-based codecs let us now concentrate our attention on a similar performance study of vector quantised (VQ) codecs.

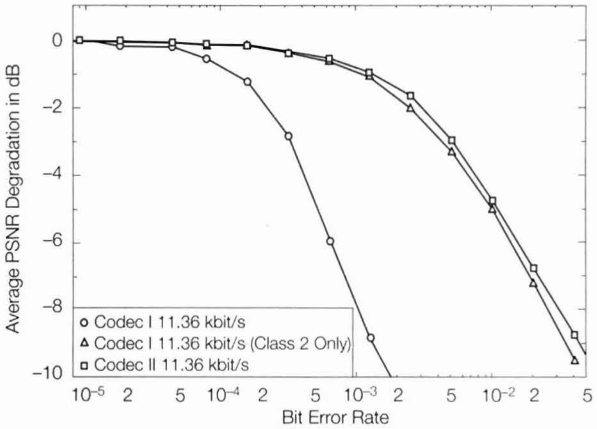


Fig. 12 - PSNR degradation versus *BER* for DCTC1 and DCTC2.

4. VECTOR-QUANTISED VIDEO CODECS

4.1. *Background*

Vector quantisation (VQ) is a generalisation of scaler quantisation, a technique lavishly documented in an excellent monograph by Gray and Gersho [55]. Extensive studies of vector quantisers have been performed by many researchers such as Huguet, Torres et al using VQ for analysis purposes [69, 70] or for image compression. In the latter field of applications finite-state VQs were studied by Yu and Venetsanopoulos [71], and Shue, Hsieh and Chuang [72] as well as Feng, Nasrabadi and Choo [73]. A variety of multi-stage, hybrid or other VQ-based arrangements were suggested by Huang et al [74] Pearlman and Lu [75], Gersho, Yip and Gupta [76],

[2] DCTC1- and DCTC2-coded sequences at various bit rates can be viewed under the WWW address http://www-mobile.ecs.soton.ac.uk or via http://www.ecs.soton.ac.uk.

Chan and Po [77], Picco and Bellifemine [78], Sin and Ngan [79], Lee and Kim [80] etc. In what follows, we will contrive a range of attractive vector quantised codecs (VQCs) [33] and comparatively study their complexity / performance equilibrium. In section 4.4 it will be demonstrated that VQCs may be rendered attractive even when using comparatively large two-dimensional (2D) vectors and small codebooks (< 512) in the codec's inter-frame mode. Specifically, we will compare the most computationally demanding full search VQ with the so-called mean-shape gain VQ, with adaptive VQ and with classified VQ. Finally we propose complexity reduction techniques and focus on bit sensitivity issues.

4.2. *VQ codec design*

The proposed VQ-based video codec shares the structure of the DCT codecs proposed in the previous section, which was shown in Fig. 7. The sole difference is that the MCER is now encoded in the time-domain using a trained codebook of 2D vectors. The specifications of the adaptive vector quantised codec (VQC) are identical to that of the adaptive DCTC, which will enable us to explore the design trade-offs associated with both implementations. Furthermore, the optional parametric coding enhancement of Fig. 7 can also be invoked.

The intra-frame mode invoked both at the commencement of communications and during the PFU operates on the basis of the same principles, as in the DCT codecs, and employs a similar bit rate control algorithm. On the same note, the *cost-gain controlled motion compensation* identifies the required number of motion-active blocks and their MVs, while for the remaining motion-passive blocks frame-differencing is invoked.

The DFD signal is now not transformed to the spatial frequency domain, instead it is passed on to the VQ scheme constituted by a trained codebook. The issues associated with the codebook training will be discussed at a later stage, in section 4.4. Now the "VQ-active" blocks providing a high coding gain are identified by an adaptive, bitrate- and cost-gain controlled algorithm and their indeces are transmitted with or without activity table compression, as we have seen for the DCT codecs. As before, the encoded frame is locally decoded and fed back to the "Previous Reconstructed Frame Buffer" in order to assist in the next stage of motion compensation. It is easily shown that for our 8 × 8 blocks 191 Floating Point Operations (Flops) are necessary for the evaluation of the *mse* associated with a particular codebook entry. In case of a 128-entry codebook for a QCIF image frame a total computational complexity of 396 blocks × 128 codebook entries × 191 Flop = 9.6 Mflops is involved, which is linearly scaled by the codebook size. In Fig. 13 we documented the achievable PSNR versus frame index performance for codebook sizes between 8 and 512 using the "Claire sequence", while fixing the bitrate at 8 kbit/s, where the best PSNR performance was exhibited by the 256-entry scheme. Note however that a similar performance was achieved by the 128-entry scheme at half the matching complexity. Using 8 address bits per 8 × 8 block the VQ data rate becomes ≈ 0.125 bits/pixel. Recall that an identical quad-class DCTC would require a total of 12 bits per 8 × 8 block for the MCER quantisation.

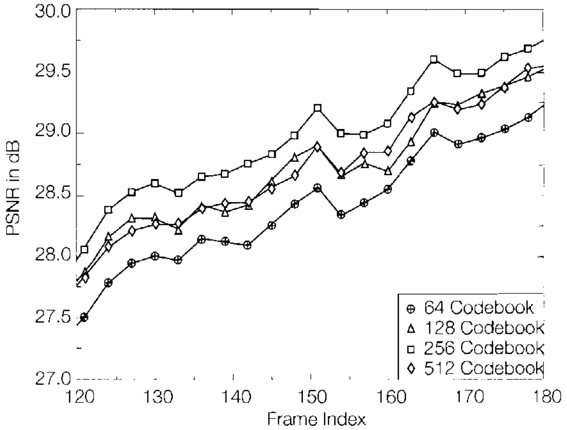


Fig. 13 - PSNR performance for the "Claire Sequence" and various codebook sizes at 10 kbit/s.

The performance of the above base-line VQC was then compared to that of a number of benchmarkers, including the so-called *mean- and variance-normalised VQ* scheme, which normalises the incoming blocks to zero mean and unity variance and uses an appropriately normalised codebook in the pattern-matching process. It can be shown that subtracting the block mean and scaling by the block variance requires $3n^2 - 2$ Flops per 8 × 8 block. When quantising and transmitting the mean and variance in order to allow the decoder to carry out the inverse operations, the codebook size is effectively increased by a factor corresponding to the product of the number of reconstruction levels of the mean and variance quantisers. Hence, a two bit quantiser for both the mean and variance would virtually expand the codebook to sixteen times it's original size. The appropriate reconstruction levels were obtained by employing a Max-Lloyd quantiser. We compared the performance of the MSVQ using VQ codebook sizes of 64 as well as 256 and Max-Lloyd mean and variance quantisers using 1 to 32 reconstructions levels, under the constraint of an 8 kbit/s constant bit rate. Using a 1-level MSVQ scheme corresponds to the baseline VQ with no scaling at all, while increasing the number of MSVQ quantisation levels at a constant bit rate inevitably reduces the number of active blocks and hence limits the video quality. Our PSNR results showed that the baseline 256-entry codec's performance was best matched by the 64-entry, 2-level scaler MSVQ scheme, which retained an approximately 0.5 dB lower PSNR at a 70% reduced codec complexity.

In an effort to further explore the complexity, quality and robustness trade-offs, in our next endeavour we implemented an *adaptive VQ* (AVQ), which was composed of a pair of buffers, one of which is termed as the active- and the other as the passive buffer. The VQ's access is limited to the active codebook, from which any of the vectors can be chosen at any instant. The most recently selected vector is fed back by the VQ to the beginning of the active codebook, but, in order to cater for less frequent vectors, after each encoding step, a system-specific number of vectors are taken from the end of the active buffer and fed-back to the passive codebook. The same number of vectors is simultaneously taken from the end of the passive codebook and entered at the beginning of the active codebook. This technique maintains a continuous vector fluctuation between the two codebooks, but allows frequently used vectors to remain in the active buffer. The increased flexibility of the method is unfortunately typically associated with an increased vulnerability against channel errors. The performance of the AVQ was evaluated again using the "Claire Sequence" and our trained 256-entry codebook, which was split into a 64-entry active buffer and a 192-entry passive buffer. The vector replenishment rate was fixed to 4, 8, 16 and 32 vectors per encoding step. The PSNR versus frame index performance shown in Fig. 14 suggests that the 64-entry VQ can approach the quality of the standard 256-entry VQ, when applying a high vector exchange rate of 32 vectors per step. At lower replenishment rates the PSNR gain over the standard 64-entry scheme was more modest.

Ramamurthi originally introduced *classified VQ* [81] for intra-frame coded images, where the encoding operation was preceeded by the classification of the block to be vector quantised. Although the classification process involves some extra computation, the classified entries are then only compared to a reduced-size codebook, which can significantly reduce the codec's complexity. In our approach a smaller codebook, derived from the same training sequence as the un-classified codebook C, can be viewed as a set of "centroids" for the codebook C. Hence, codebook C can be split into n codebooks c_n by assigning each vector of the original codebook C to one of the centroids. The encoding procedure then consists of a two-stage VQ process and may be viewed as a so-called tree-structured VQ (TSVQ) [81]. The classifier is a VQ in its own right, using a codebook filled with the n centroids.

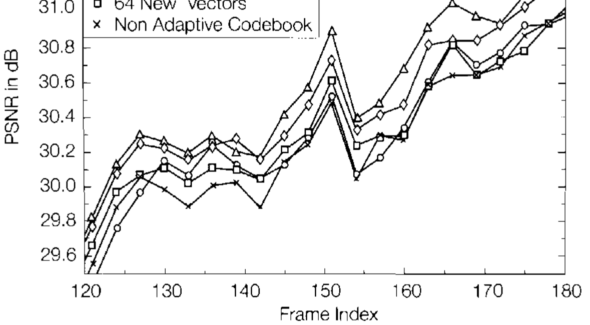


Fig. 14 - PSNR versus frame index performance for the "Claire" sequence with and without adaptive codebooks at 10 kbit/s.

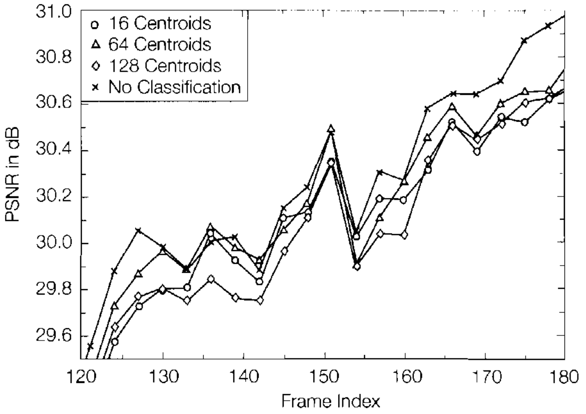


Fig. 15 - PSNR versus frame index performance of classified VQ codecs when using the "Claire" sequence at 10 kbit/s and a 256-entry codebook.

Once the closest centroid has been found, the sub-codebook containing the associated vectors is selected for the second VQ step. The performance of the CVQ was evaluated, based on the optimum 256-entry codebook, with classifiers containing n = 16, 64 and 128 centroids, as portrayed in Fig. 15. The PSNR performance loss due to using this sub-optimum two-stage approach is less than 0.3 dB, while the number of block comparisons was reduced from 256 to around 64 for the CVQ.

4.3. *VQC1 and VQC2 performance*

Similarly to the previously proposed DCT-based codecs, we contrived two VQ codecs, VQC1 and VQC2. VQC1 achieved a higher compression ratio due to using the previously proposed table compression algorithms, while VQC2 exhibited a higher innate robustness against channel errors. Both codecs are based on the classified VQ codec, which offered the best compromise in terms of quality, robustness and complexity. We used a centroid codebook size of 16 and an overall codebook size of 256 which lead to an overall codec complexity of around 15 Mflops, when employing the active / passive classification. The MCER was generated for all 396 8 ×

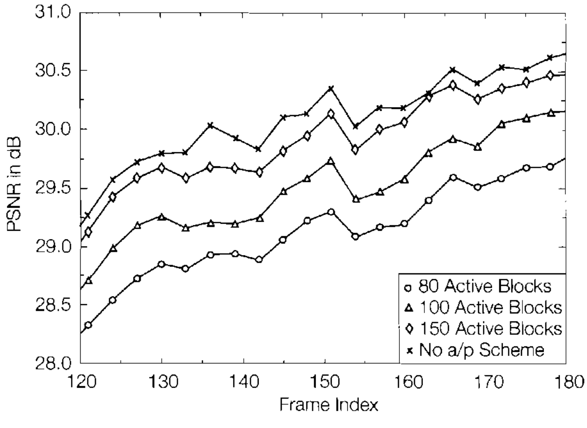


Fig. 16 - PSNR degradation versus frame index performance of active/passive classified (16 centroids/ 256 codebook size) VQ codecs for the "Claire" sequence at 10 kbit/s.

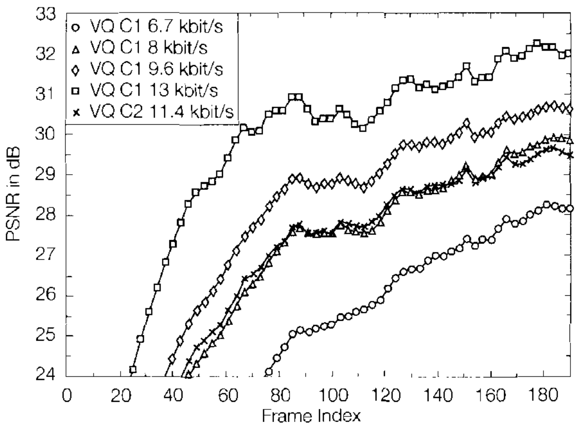


Fig. 17 - PSNR versus frame index performance of VQC1 at various bit rates and for VQC2 at 11.3 kbit/s for the "Claire" sequence.

8 blocks and a bit-rate constrained fraction of the highest-energy 20-50% MCER blocks were vector quantised. The PSNR penalty of limiting the proportion "VQ-active" blocks is quantified by Fig. 16 using the "Claire" sequence for active block fractions of 20-50%, when invoked in the 8 kbit/s VQC1 scheme.

The peak signal-to-noise ratio (PSNR) versus frame index performance of the VQC1 scheme is portrayed in Fig. 17 for the "Claire" sequence at the previously introduced 2-nd generation mobile radio speech bit rates of 6.7, 8, 9.6 and 13 kbit/s. Lastly, the associated bit allocation schemes are summarised in Table 4 in contrast to our other prototype codecs ([3]).

then improved in a second stage using the so-called generalised Max-Lloyd algorithm [14]. The VQ's training sequence has to be statistically representative of a wide range of images. In order to arrive at an adequate training set we generated the MCER signal for a range of input sequences, while constraining the number of active MVs to 35 out of the 396. This constraint inherently assumed a low-rate application. Then the 35 highest-energy MCER blocks of each video frame were included in the training set. The Probability Density Function (PDF) of the MCER's mean and variance is portrayed in Fig. 18. The mean and variance are highly peaked between |-2 ... 2| and [20 ... 40], respectively, which demonstrates the efficiency of MC, since in its mean tends to be close to zero and its variance implies a low MCER energy.

In our first attempt we designed the required 256-entry codebook using the so-called "pruning" method [55], where the codebook is initialised with an arbitrary vector and then further training set vector are included in the codebook, if they are deemed sufficiently dissimilar in order to statistically represent another ensemble of MCER blocks. However, this technique yielded a codebook, which exhibited an inadequate video performance, even when the codebook size was increased to 512 entries. We then endeavoured to improve its performance by using the so-called "pairwise nearest neighbour" (PNN) technique [55] summarised in Algorithm 1, which reduces the initial training set to the required codebook size step by step, until a desired size is achieved.

Table 4 - *Bit allocation table*

Codec	FAW	PFU	MV Index + MV	DCT Ind. + DCT	VQ Ind. + VQ	QT + PC	Padding	Total
DCTC2	22	22 × 4	30 × 9 + 30 × 4	30 × 9 + 30 × 12	-	-	6	1136
DCTC1	22	22 × 4	< 350 (VLC)	< 350 (VLC)	-	-	VLC	800
CQC2	22	22 × 4	38 × 9 + 38 × 4	-	31 × 4 + 31 × 8	-	5	1136
VQC1	22	22 × 4	< 350 (VLC)	-	< 350 (VLC)	-	VLC	800
QTC1	22	22 × 4	< 500 (VLC)	-	-	< 565 + 1 or 80	VLC	1136

During our previous discussions appropriately trained codebooks were used and in the next section the training procedures employed will be briefly highlighted.

4.4. *Codebook design*

Our proposed VQ codebook design algorithm is similar to the one used in the DCT codec's quantiser training in that an initial codebook is derived first, which is

Algorithm 1 *The pairwise nearest neighbour algorithm summarised below reduces the initial training set to the required codebook size step by step, until a desired size is achieved.*

1) Assign each training vector to a so-called cluster of vectors.
2) Evaluate the potential distortion penalty associated with tentatively merging each possible pairs of clusters.
3) Carry out the actual merging of that specific pair of clusters, which inflicts the lowest excess distortion due to merging.

[3] VQC1 and VQC2 encoded sequences at various bit rates can be viewed under the WWW address http://www-mobile.ecs.soton.ac.uk.

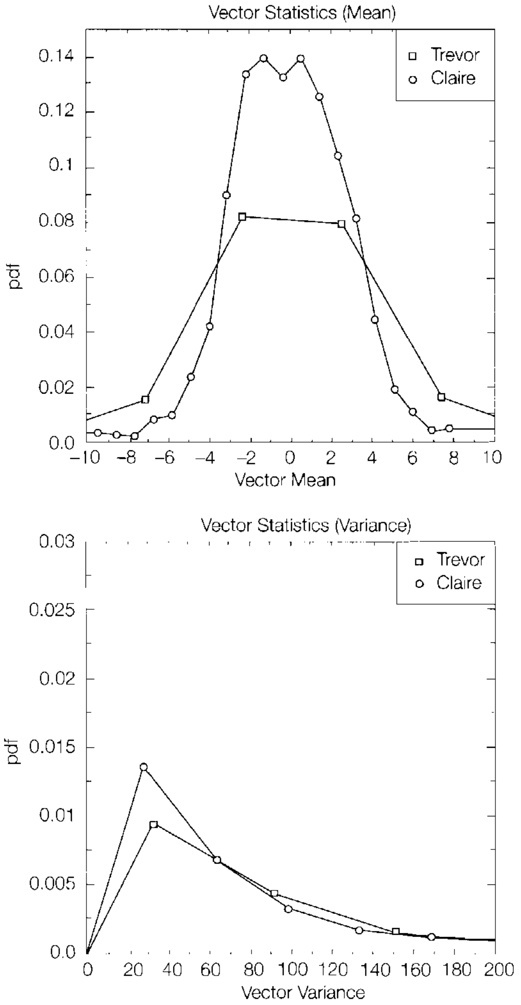


Fig. 18 - Statistical properties of some VQ training sequences in terms of mean and variance (Claire / Trevor).

4) Repeat Steps 2/ and 3/, until the required codeboook size is arrived at.

The algorithm's complexity became excessive for our large training set, therefore we simplified this algorithm in a sub-optimum fashion. Namely, instead of merging each possible pair of clusters, where the number of combinations exhibits a quadratic expansion with increasing codebook size, we preselected a single cluster, which was combined with all the others. This sub-optimum algorithm was initially invoked to reduce the training set and then the full PNN algorithm was invoked in order to create a range of codebooks having sizes between 4 and 1024.

In the second step of the codebook generation we used the generalised Max-Lloyd algorithm to enhance our initial codebooks, which resulted in a limited reduction of the average codebook distortion. This underlines the fact that our initial codebooks generated by the PNN algorithm were adequate and the suboptimum two-stage approach caused only a negligible performance loss. As an example, our 128-entry codebook constituted by 8 × 8 vectors is depicted in Fig. 19, where the vector components were shifted by 127 and multiplied by 30 in order to visually emphasize their differences.

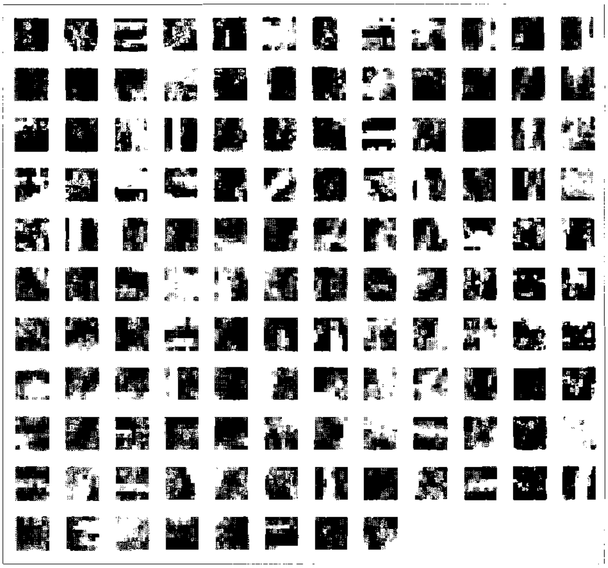


Fig. 19 - Enhanced sample codebook with 128 8 × 8 vectors.

Here we refrain from elaborating on the robustness issues of VQC1 and VQC2 due to lack of space. Their fundamental behaviour under erroneous channel conditions [33] is akin to that of DCTC1 and DCTC2, respectively, which was shown in Fig. 12. These issues will be comparatively studied for DCT, VQ and QT codecs in our Conclusions Section. Following the design and performance evaluation of VQ codecs let us now consider quad-tree (QT) coded schemes.

5. Quad-tree based codecs

5.1. *Background*

Quad-tree (QT) codecs were studied by a number of authors, including Strobach [57], Vaisey and Gersho [58], Shustermann and Feder [59], Arnold et al [82] etc. These schemes also obey the structure of Fig. 7.

5.2. *Quadtree decomposition*

In case of the DCT and VQ codecs the MCER was split in perfectly tiling blocks and we employed the active / passive block classification scheme, which was then eventually described by the help of an activity table, in order to adequately model the MCER. QT-decomposition constitutes an alternative way of describing the structure of the MCER, where instead of using a regular block structure, the MCER is decomposed in variable size sectors characterized by similar features, for example by similar grey levels. Explicitly, the MCER residual frame portrayed in the centre of Fig. 21 is described in terms of two sets of parameters, the structure or spatial distribution of similar regions and their grey levels. Again, similarly to the activity tables in the proposed DCT and VQ codecs, the QT structure is much more sensitive to bit errors than the grey level coding bits.

An example of QT decomposition is portrayed in Fig. 20 in a so-called top-down approach, where the MCER residual was initially split into four quadrants, since its texture was too inhomogeneous and hence it did not satisfy the so-called "similarity criterion", which would allow treating it as a homogeneous domain. When any MCER section does satisfy the appropriately adjusted, quality-dependent similarity criterion, it can be treated as a homogeneous grey area, as it is examplified by the top left and bottom right corners of Fig. 20. The two other quadrants of the Figure are subjected to a number of further decomposition steps until the similarity criterion is satisfied or the bitrate budget is exhausted. Clearly, the irregular QT structure constitutes an alternative way of describing the previously introduced DCT- and VQ-activity table information of the DCT and VQ codecs, although for the QT codec a finer spatial MCER decomposition must be maintained, since the irregularly decomposed blocks are now simply modelled by their mean values. The derivation of the variable-length QT-structure description code shown at the bottom of Fig. 20 now becomes straightforward. Each inhomogeneous so-called "parent node" is flagged with a binary one classifier, while the homogeneous segments - also often referred to as "leaf nodes" - are denoted by a binary zero classifier and the flags are read from top to bottom and left to right. It can be inferred from Fig. 20 that the location and size of 13 different segments can be encoded using a total of 17 bits. The reconstructed frame is shown at the right hand side of Fig. 20, whereas the typical segmentation of a frame is exemplified in Fig. 21, where the QT structure is portrayed with and without the overlaid MCER residual frame and the original video frame. In the eye and lip regions a more stringent similarity match was required than in the background, which led to a finer QT decomposition. Another design alternative is to use a more coarse QT decomposition of the MCER and employ a more sophisticated, higher resolution QT block description, which will be addressed in the next section.

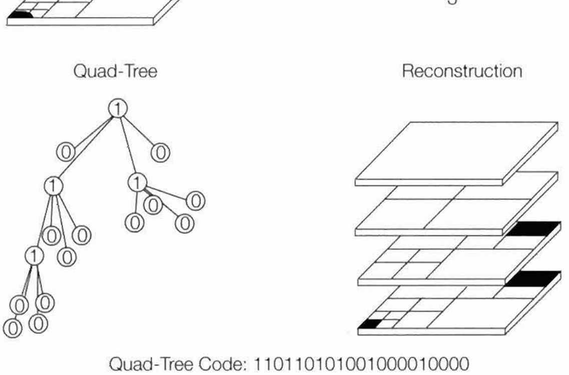


Fig. 20 - Regular decomposition / reconstruction example and the corresponding quadtree.

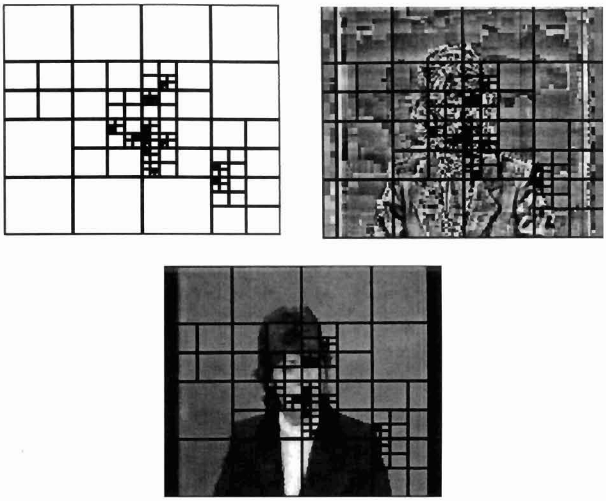


Fig. 21 - Quad-tree segmentation example with and without overlaid MCER residual and original video frame.

5.3. *Quadtree block intensity match*

As mentioned above, in case of coarse QT decomposition, or if high image quality requirements must be met, the simple block average representation of the QT decomposed blocks can be substituted by more sophisticated techniques, such as vector quantisation (VQ), discrete cosine transformed (DCT) [65] or subband (SB) decomposed representations. However, a fine QT decomposition requires applying VQ or DCT to blocks of different sizes, which increases the codec's complexity. Instead, in order to explore the range of design trade-offs we first studied the performance of a low-complexity zero-order mean- or average-intensity model, which was then compared in performance terms to a first-oder luminance intensity profile.

5.3.1. ZERO-ORDER BLOCK INTENSITY MATCH

In order to design the appropriate mean intensity quantisers, we evaluated the PDF of the average values m of the variable sized MCER blocks. As expected, for various QT-decomposed MCER block sizes quantisers having different mean values and variances were required, since the mean of the MCER blocks towards the top of the QT, namely at QT Levels 2-4, which cover a large area is more likely to be close to zero than the mean of smaller blocks, which exhibited itself in a more highly peaked and hence less spread PDF. The mean of the smallest blocks at QT Levels 5-7 tends to fluctuate over a wide range, yielding a near-uniform PDF for Level 7. Hence

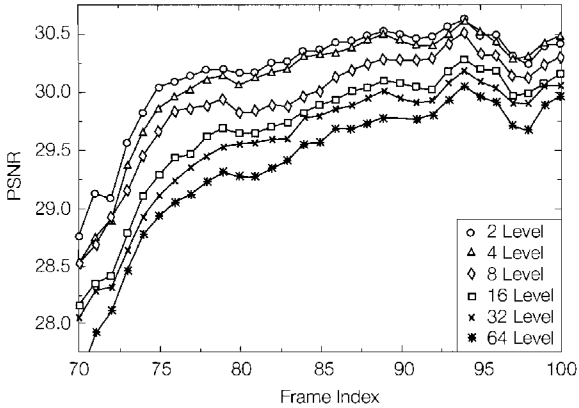


Fig. 22 - PSNR versus frame-index performance of the QT codec for various quantisers at a constant bit rate of 10 kbit/s.

we designed appropriate Max-Lloyd quantisers for each QT hierarchy level. Similarly to our previous approach as regards to DCT and VQ coding, a two-stage quantiser training procedure was contrived. First the unquantised mean values for each QT-decomposed block size were recorded using a training sequence in order to derive an initial set of quantisers. During the second, true training stage this initial set of quantisers was then used tentatively in the QT codec's operation in order to record future unquantised block averages generated by the codec operated at a rate of 1000 bits/frame, which was the average rate for the targeted 2-nd generation mobile videophone systems as well as for the adaptive IMT. This two-stage approach was necessary to obtain realistic training data for the Max-Lloyd training of the final quantisers.

We then generated codebooks for a range of different number of reconstruction levels and endeavoured to determine the best distribution of coding bits between the QT-code and block intensity encoding. Fig. 22 characterises the codec's performance for various quantisers ranging from two to 64 level schemes, while generating 1000 bits/frame. Observe in the Figure that the two-level and four-level quantisers were found to have the best performance, since the adaptive bit rate control algorithm restricted the number of hierarchy levels in the QT decomposition process, when more bits were allocated to quantise the block averages. Therefore it was disadvantageous in terms of both objective and subjective quality to finely resolve the block luminance intensity at the cost of reducing the QT resolution. A more efficient exploitation of the higher number of QT block description bits may be in this case to have a more elaborate QT block intensity model, such as the first-order model or the previously mentioned DCT-, VQ- or SBC-based QT block representation, if a higher complexity is acceptable.

5.3.2. First-order intensity match

In order to verify this hypothesis we then embarked on studying the performance of the first-order luminance-intensity surface defined by

$$b\,(x, y) = a_0 + a_{x1}\,x + a_{y1}\,y \tag{3}$$

where, in order to describe the luminance $b\,(x, y)$ of the pixel at coordinates x and y, the coefficients a_{x1} and a_{y1} must be known. In this case the block's luminance is approximated by a plane sloping in both x and y directions. The squared error of this linear approximation is given by

$$e = \sum_{x=1}^{b}\sum_{y=1}^{b}\left[b(x,y) - \left(a_0 + a_{x1}\,x + a_{y1}\,y\right)\right]^2 \tag{4}$$

where the constants a_0, a_{x1} and a_{y1} are determined by setting the partial derivative of eq. (4) with respect to all three variables to zero and solving the resulting three dimensional problem. In order to assess the performance of this scheme Max-Lloyd quantisers were designed for each of the constants and the PSNR performance of a range of quantisers using different number of quantisation levels was tested. We found that the increased number of quantisation bits required for the three different coefficient quantisers was too high to facilitate a sufficiently fine QT-based MCER residual decomposition. Hence in our targeted bitrate region around 10 kbit/s the PSNR performance of the first order scheme became by about 0.5 dB inferior to that of the zero-order model under our specific bit rate constraints. Therefore we favoured the zero-order model.

5.4. *QT decomposition algorithmic issues*

Most recently proposed QT codecs [57], [82] invoke the previously outlined thresholding process in order to decide whether a tree should be further decomposed. Due to the typically time-variant nature of the video sequences processed, the threshold-based QT decomposition usually yields an approximately time-invariant video quality associated with a fluctuating, time-variant bit rate. However, our proposed adaptive videophone codec is required to guarantee a constant, but programmable bit rate. A further requirement in very low bitrate video communications is gain-cost control, which is difficult to achieve in the framework of QT schemes, since the gain of a potential decomposition becomes known only after the decomposition took place.

Having assessed the potential of a number of different approaches to contriving an appropriate adaptive bit allocation scheme we finally arrived at Algorithm 2 [32]. Accordingly, the codec develops the QT structure down to a given maximum number of decomposition levels and then determines the gain of each decomposition step by evaluating the difference between the mean squared video reconstruction error of the "parent block" and the total *mse* contribution associated with the sum of its four "child blocks". Exposition of the Algorithm is further aided by referring to Fig. 20.

Algorithm 2 *This algorithm adaptively adjusts the required QT resolution, the number of QT description*

bits and the number of encoding bits required in order to arrive at the target bit rate [32].

1) Develop the full tree from minimum to maximum number of QT levels (eg 2-7).
2) Determine the *mse* gains associated with all decomposition steps for the full QT.
3) Determine the average decomposition gain over the full set of leaves.
4) If the potentially required number of coding bits is more than twice the target number of bits for the frame, then delete all leaves having less than average gains and repeat Step 3.
5) Otherwise delete leaves on an individual basis, starting with the lowest gain leaf, until the required number of bits is attained.

Observe that the purpose of Steps 3 and 4 is to introduce a bitrate-adaptive, computationally efficient way of pruning the QT to the required resolution. This allows us to incorporate an element of cost-gain quantised coding, while arriving at the required target bit rate without many times tentatively decomposing the image in various ways in an attempt to find the optimum fixed bit allocation scheme. The algorithm typically encountered 4-5 such fast QT pruning recursions, before branching out to Step 5, which facilitated a slower converging fine-tuning phase during the bit allocation optimisation.

Viewing from a different angle, Algorithm 2 deleted those leaves from the tree that resulted in the lowest decomposition gains. During this QT pruning process the elimination of leaves converted some of the nodes to leaves, which were then considered for potential elimination during future coding steps. Therefore the list of leaves associated with the lowest decomposition gains had to be updated before each QT pruning step. During the fast pruning phase constituted by Steps 3 and 4 in a computationally efficient, but sub-optimum approach, we determined the average gain of the entire set of leaves and rather than deleting the leaves associated with the lowest decomposition gain one-by-one, all leaves that had a gain lower than the average gain were deleted in a single step. The slower one-by-one pruning constituted by Step 5 was then invoked before concluding the bit allocation in order to fine-tune the number of bits allocated. This sub-optimum deletion process was repeated until the tree was pruned to the required size and the targeted number of coding bits was allocated.

In Fig. 23 we displayed both the maximum possible and the recorded average number of nodes associated with each tree level, where a logarithmic vertical axis was used in order to allow us to accommodate a range of values between 1 and 16384 nodes / leaves. Although it is not accurately readable on the logarithmic axis, during our experiments we noted that 15.3/16 = 95.6% of the time the QT depth exceeded three levels, which prompted us to prune the minimum tree depth to three without sacrificing coding efficiency. We also observed that only about 52.9/16 384 = 0.32% of the potentially possible Level 7 leaves was ever decomposed and therefore we limited the maximum tree depth to six. Albeit the codec's complexity was reduced, this was associated with a concomitant average PSNR reduction of around 1 dB. Hence, if video quality was at premium, the 7-th QT level was necessary in order to develop important fine details in the frame.

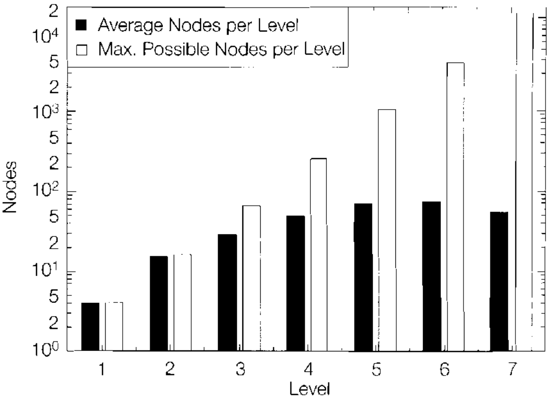


Fig. 23 - Average tree nodes and leaves per hierarchical level within the tree at 1200 bpf and maximum possible tree nodes.

In summary of our QT-coding investigations we concluded that due to the inherent error sensitivity of the QT-description code all codecs are of Type 1 and their compression ratio is slightly more modest than that of the similar-robustness equivalent Type 1 DCTC1 and VQC1 schemes. Viewed from a different angle, the Type 1 QT codecs exhibit similar bitrates to the more robust Type 2 DCTC2 or VQC2 arrangements. The PSNR versus frame index performance of the best zero-order codecs was shown in Fig. 22 (4), while its robustness evaluated in terms of PSNR versus *BER* will be compared to that of our other benchmarkers in Fig. 27. Having described the proposed adaptive QT codec let us now focus our attention on the performance comparison of the various proposed schemes.

6. Conclusions

In section 2 we compared a range of MC techniques and analysed their efficiency in the light of the bit rate investment required for the MVs. We found that the best design trade-offs were maintained, when using cost/gain controlled MC. Section 3 capitalised on these results in order to contrive fully cost/gain controlled DCT based coding schemes, which exhibited a constant, but programmable bitrate. Two different schemes, the more bandwidth efficient, but more error sensitive DCTC1 and the slightly higher rate but more robust DCTC2 were proposed for various applications. The objective of section 4

(4) Examples of QT-coded sequences can be viewed under the following WWW address: http:/www-mobile.ecs.soton.ac.uk.

was to derive a set of VQ codecs, in particular VQC1 and VQC2. In section 5 we presented a range of quad-tree based schemes. All proposed codecs have also been subjected to bit sensitivity analysis in order to assess their performance under erroneous conditions and to assist in the design of source matched FEC schemes for various channels. Some of these findings will be reviewed in this section. The associated transmission issues, including source sensitivity-matched forward error correction (FEC) coding, reconfigurable modulation and automatic repeat request (ARQ) schemes have been discussed in depth in a series of companion papers published by the authors, namely in [31, 32, 33], respectively.

Let us finally compare our proposed inter-frame codecs to standard codecs, such as the MPEG-2, the H261 and the H263 codecs. The latter standard schemes are typically variable rate codecs, which make extensive use of variable-length compression techniques, such as run-length (RL) coding and entropy coding [14], although it is possible to invoke appropriate adaptive packetisation and multiple consecutive encoding operations in order to arrive at a required near-constant bitrate. This philosophy was followed Cherriman and Hanzo in [83, 84, 85]. An often employed alternative solution in distributive video applications is to use a buffer with a feed-back to the quantiser, instructing the codec to invoke more coarse quantisers, when the buffer fullness exceeds a certain critical limit. Using buffering in interactive videotelephony is not a realistic alternative, since in case of 10 frames/s scanning the inherent latency is 100 ms without any buffering. In case of buffering this would be increased. The voice signal's latency or delay is becoming annoying for delays of 100 ms, hence in case of buffered video perfect lip-synchronisation cannot be realistically achieved.

The above standard codecs also require the transmission of at least one intra-coded (*I*) frame at the commencement of transmission in order to provide a reference for the operation of the motion compensation. The transmission of the so-called *I* frames can be repeated at selectable regular intervals, in order to replenish the reconstructed frame buffer of the decoder, thereby mitigating the effect of prolonged transmission errors, yielding a regular surge in the bit rate. This is unacceptable in conventional fixed-rate mobile radio systems. In distributive video systems these bitrate surges are smoothed by the adaptive buffers at the cost of a slight delay. Furthermore, if the *I*-frame is corrupted, it may inflict more severe video degradation than that due to previous inter-coded frame errors. In addition to the I frames, the H.263 and MPEG-2 codecs use two more modes of operation, namely, the so-called predicted coding (*P*) and bi-directional (*B*) coding modes, where the *P*-frames rely on differential coding strategies invoked with reference to the surrounding *I* and *B* frames. Due to the above robustness and delay problems we found that our distributed partial forced update (PFU) scheme was more appropriate for the targeted mobile radio applications. The corresponding design trade-offs of H.263-based candidate systems were studied in [83, 84, 85], where the variable-rate, run-length coded H.263 scheme's robustness was improved by Automatic Repeat Request (ARQ) assistance, while the bitrate fluctuation was mitigated by adaptive packetisation.

In our experiments portrayed in Fig. 24 we stipulated a fixed bit rate of 10 kbit/s for our three prototype codecs and adjusted the parameters of the H261 and MPEG-2 codecs to provide a similar video quality associated with a similar average PSNR performance of around 30 dB. The PSNR versus bitrate performance of the proposed codecs is shown in Fig. 25. The corresponding PSNR curves are displayed in Fig. 26. Observe in Fig. 24 that the number of bits / frame for our proposed codecs is always 1000, corresponding to 10 kbit/s and it is about twice as high for the two standard codecs, exhibiting a random fluctuation for the H261 codec. The MPEG codec exhibits three different characteristic bit rates, corresponding to the *I*, *B* and *P* frames in decreasing order from around 8000 bits / frame, to about 1800 and 1300, respectively. Hence, due to their variable bitrates the above standard video codecs are not amenable to employment in interactive videophony over existing fixed-rate mobile systems.

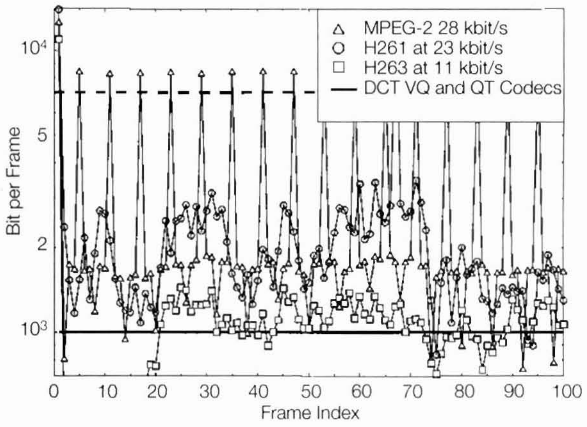


Fig. 24 - Bit rate fluctuation versus frame index for the proposed adaptive codecs and two standard codecs.

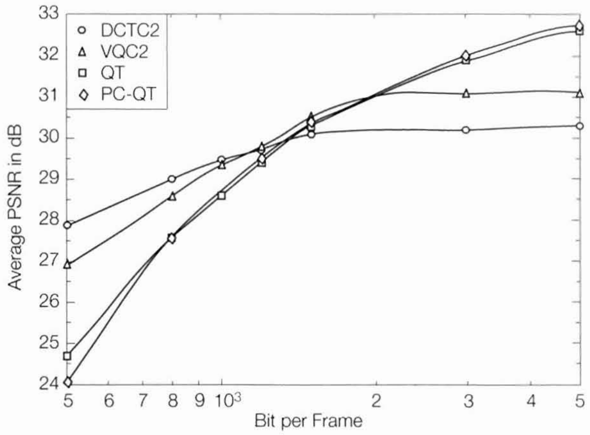


Fig. 25 - PSNR versus bit rate performance of the proposed adaptive codecs.

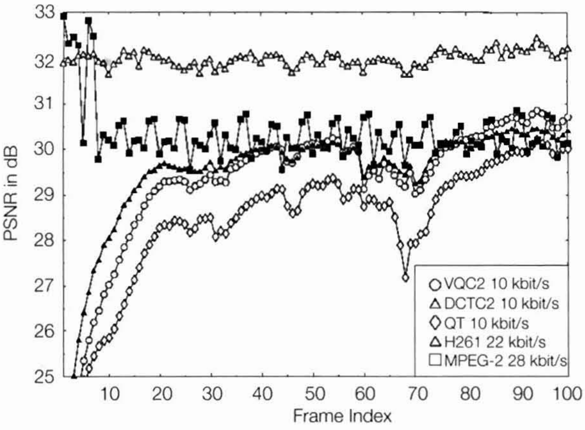


Fig. 26 - PSNR versus frame index performance of the proposed adaptive codecs and three standard codecs.

The delay of our codecs and that of the H-261 codec is in principle limited to one frame only. The delay of the H.263 and MPEG-2 codecs may stretch to several frames due to the *P*-frames. In order to smoothe the teletraffic demand fluctuation of the MPEG-2 codec typically adaptive feedback controlled output buffering is used, which further increases the delay.

The error resilience of the Type 1 codecs - namely that of DCTC1, VQC1 and QT - which use the run-length-compressed active / passive table concept is very limited, as is that of the standard codecs. These arrangements have to invoke Automatic Repeat Request (ARQ) assistance over error-prone channels. Hence in these codecs single bit errors can corrupt an entire frame, or in fact several frames in case of the MPEG-2 codec. These problems are avoided by the slightly less bandwidth efficient non-run-length encoded Type 2 schemes, which therefore exhibit an improved error resilience.

Overall, the vector quantised codecs VQC1 and VQC2 constitute the best compromise in terms of quality, compression ratio and computational demand, closely followed by the DCTC1 and DCTC2 candidate codecs. In our concluding comparative test we input the artificial test pattern depicted in the top left corner of Fig. 28 to the DCT, VQ and QT inter-frame codecs and portrayed the evolution of the reconstruction error's PDF as a function of the frame index. The reconstruction error's PDF is spread over the widest range in case of the QT codec, which gradually clusters around smaller error magnitudes, as time elapses, but it never reaches the lower error variance of the DCT or VQ codecs. Furthermore, the QT codec does not lend itself to Type 2 implementations, since the QT code is rather vulnerable to channel errors, as it is evidenced by Fig. 27.

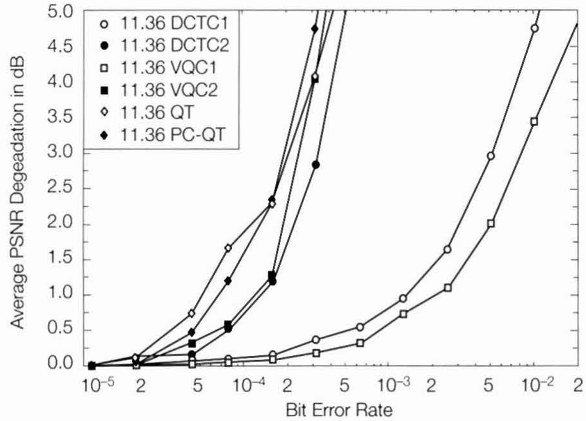


Fig. 27 - PSNR degradation versus *BER* for the proposed codecs.

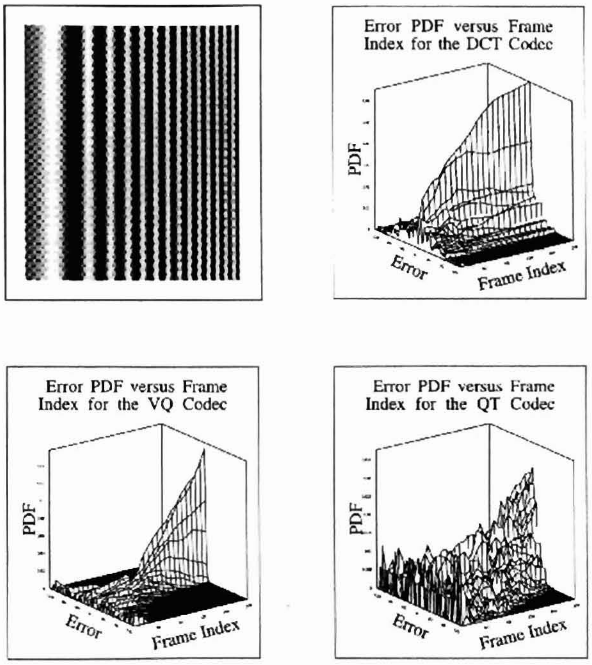


Fig. 28 - Error PDF for the Test-Pattern depending on the frame index.

In conclusion, a range of gain-cost quantised, fixed-rate videophone codecs has been proposed and studied comparatively for employment in fixed-rate interactive wireless systems, where the error-sensitive, variable-rate standard codecs cannot be employed. The best candidate codec was VQC2. In our current work we embarked on a similar comparative study of higher resolution, namely CIF, 4 × -CIF and 16 × -CIF based robust, programmable-rate schemes for wireless local area network (WLAN) applications.

Acknowledgement

The financial support of the EPSRC, UK in the framework of the research contract GR/K74043 is gratefully acknowledged. Furthermore, the authors would like to express their thanks to the anonymous Reviewers and to the numerous authors referenced whose ideas contributed towards our work.

Manuscript received on May 2, 1996.

References

[1] L. Hanzo, J. Stefanov: *Mobile radio communications.* ch.The Pan-European Digital Cellular Mobile Radio System known as GSM. Pentech Press, London, 1994, p. 677-773.

[2] Telcomm. Industry Association (TIA), Washington: *DC, Dual-mode subscriber equipment - Network equipment compatibility specification.* Interim Standard IS-54, 1989.

[3] Telcomm. Industry Association (TIA), Washington: *DC, Mobile station - Base station compatibility standard for dual-mode wideband spread spectrum cellular system.* EIA/TIA Interim Standard IS-95, 1993.

[4] Research and Development Centre for Radio Systems, Japan: *Public Digital Cellula*r (PDC) Standard, RCR STD-27.

[5] International Standards Organization: *ISO/IEC 11172 MPEG 1 International Standard, Coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s, Parts 1-3.*

[6] International Standards Organization: *ISO/IEC CD 13818 MPEG 2 International Standard, Information Technology, Generic coding of moving video and associated audio information, Parts 1-3.*

[7] D. Greenwood, L. Hanzo: *Mobile radio communications.* ch. Mobile Radio Channels. London: IEEE Press-Pentech Press, 1992, p. 92-185.

[8] L. Hanzo, J. Woodard: *A dual-rate 4.8/7.2 kbit/s algebraic celp-based speech transceiver.* IEEE Conference on Vehicular Technology (VTC), June 1994, p. 1690-1694.

[9] L. Hanzo, J. Woodard: *An intelligent voice communications system for indoors communications.* IEEE Conference on Vehicular Technology (VTC), August 1995.

[10] J. Williams, L. Hanzo, R. Steele, J. Cheung: *A comparative study of microcellular speech transmission schemes.* "IEEE Transactions on Veh. Technology", Vol. 43, November 1994, p. 909-925.

[11] L. Hanzo, J. Streit, R.A. Salami, W.T. Webb: *A low-rate multi-level voice/video transceiver for personal communications.* "Wireless Personal Communications", Vol. 2, No. 3, 1995, p. 217-234.

[12] L.Hanzo, J.C.S. Cheung, R. Steele, W.T. Webb: *A packet reservation multiple access assisted cordless telecommunications scheme.* "IEEE Transactions on Veh. Technology", Vol. 43, May 1994, p. 234-245.

[13] A. Netravali, B. Haskell: *Digital pictures: representation and compression.* Plenum Press, 1988.

[14] A. K. Jain: *Fundamentals of digital image processing.* Prentice-Hall, 1989.

[15] N. Jayant, P. Noll: *Digital coding of waveforms, Principles and applications to speech and video.* Prentice Hall, 1984.

[16] K. H. Tzou, H. G. Mussmann, K. Aizawa, (Ed.): *Special issue on very low bit rate video coding.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 4, June 1994, p. 213-357.

[17] N. Hubing (Ed.): *Speech and image coding.* "Special Issue of the IEEE Journal on JSAC", Vol. 10, June 1992, p. 793-976.

[18] B. Girod et al (Ed.): *Special issue on image sequence compression.* "IEEE Transactions on Image Processing", Vol. 3, September 1994, p. 465-716.

[19] M. Khansari, A. Jalali, E. Dubois, P. Mermelstein: *Robust low bit-rate video transmission over wireless access systems.* Proc. of International Comms. Conf. (ICC), 1994, p. 571-575.

[20] R. M. Pelz: *An unequal error protected px8 kbit/s video transmission for dect.* Vehicular Technology Conference, IEEE, 1994, p. 1020-1024.

[21] T. Chen: *A real-time software based end-to-end wireless visual communications simulation platform.* Proc. of SPIE Conf. on Visual Communications and Image Processing, 1995, p. 1068-1074.

[22] K. Illgner, D. Lappe: *Mobile multimedia communications in a universal telecommunications network.* Proc. of SPIE Conf. on Visual Communications and Image Processing, 1995, p. 1034-1043.

[23] Y. Zhang: *Very low bit rate video coding standards.* Proc. of SPIE Conf. on Visual Communications and Image Processing, 1995, p. 1016-1023.

[24] H. Ibaraki et al: *Mobile video communication techniques and services.* Proc. of SPIE Conf. on Visual Communications and Image Processing, 1995, p. 1024-1033.

[25] K. Watanabe et al: *A study on transmission of low bit-rate coded video over radio links.* Proc. of SPIE Conf. on Visual Communications and Image Processing, 1995, p. 1025-1029.

[26] P. Sheldon, J. Cosmas, A. Permain: *Dynamically adaptive control system for mpeg-4.* Proceedings of the 2-nd International Workshop on Mobile Multimedia Communications, 1995.

[27] ITU-T: *Draft Recommendation H.263: Video coding for low bitrate communication.* Due to be approved 1995/96.

[28] Telenor Research and Development, P.O.Box 83, N-2007 Kjeller, Norway, H.263 Software Codec.

[29] ITU-T: *Recommendation H.261: Video codec for audiovisual services at p × 64 kbit/s.* March 1993.

[30] P. Cherriman, L. Hanzo, R. Lucas: *ARQ-assisted h261 and h263-based programable-rate video transceivers.* "Research Journal of the Dept. of Electr. and Comp. Sc.", 1995.

[31] L. Hanzo, J. Streit: *Adaptive low-rate wireless videophone schemes.* "IEEE Transactions Video Technology", Vol. 5, August 1995, p. 305-319.

[32] J. Streit, L. Hanzo: *Quad-tree based reconfigurable videophone systems.* Vol. 6, No. 2, April 1996, p. 225-237.

[33] J. Streit, L. Hanzo: *Vector-quantised low-rate cordless videophone systems.* IEEE Transactions Vehicular Technology, Vol. 42, No. 2, Mag 1997, p. 340-351.

[34] L. Hanzo, R. Stedman, R. Steele, J. Cheung: *Multi-media technologies and future applications, ch.* A portable multi-media communicator scheme. Pentech Press, 1994, p. 31-54.

[35] J. Streit, L. Hanzo: *A fractal video communicator.* IEEE Conference on Vehicular Technology, (Stockholm, Sweden), June 1994, p. 1030-1034.

[36] B. Liu, A. Zaccarin: *New fast algorithms for the estimation of block motion vectors.* "IEEE Transactions on Circuits and Systems", Vol. 3, April 1993, p. 148-157.

[37] R. Li, B. Zeng, N. Liou: *A new three step search algorithm for motion estimation.* "IEEE Transactions on Circuits and Systems", Vol. 4, August 1994, p. 439-442.

[38] G. Bozdagi et al: *3-d motion estimation and wireframe adaptation including photometric effects for model-based coding of facial image sequences.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 4, June 1994, p. 246-256.

[39] L. Lee, J. Wang, J. Lee, J. Shie: *Dynamic search-window adjustment and interlaced search for block-matching algorithm.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 3, February 1993, p. 85-87.

[40] C. Huang, C.Hsu: *A new motion compensation method for image sequence coding using hierarchical grid interpolation.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 4, February 1994, p. 42-51.

[41] J. Nieweglowski, T. Moisala, P. Haavisto: *Motion compensated video sequence iinterpolation using digital image warping.* "IEEE ICASSP", 1994, p. 5-205 - 5-208.

[42] B. Girod: *Motion-compensating prediction with fractional-pel accuracy.* "IEEE Transactions on Communications", Vol. 41, April 1993, p. 604-611.

[43] J. Huang et al: *A multi-frame pel-recursive algorithm for varying frame-to-frame displacement estimation.* "IEEE ICASSP-92", 1992, p. III-241 - III-244.

[44] N. Efstratiadis, A. Katsaggelos: *Adaptive multiple-input pel-recursive displacement estimation.* "IEEE ICASSP-92", 1992, p. III-245 - III-248.

[45] J. Jain, A. Jain: *Displacement measurement ans its applications in inter frame image coding.* "IEEE Transactions on Communications", Vol. 29, December 1981.

[46] B.-M. Wang et al: *Zero waiting-cycle hierarchical block matching algorithm and its array architectures.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 4, February 1994, p. 18-27.

[47] J. Woods, S. O'Neil: *Subband coding of images.* "IEEE Trans. ASSP", Vol. 34, October 1986, p. 1278-1288.

[48] J. Woods: *Subband image coding.* Kluwer Academic Publishers, 1991.

[49] H. Gharavi: *Subband image coding, ch.* Subband Coding of Video Signals, Kluwer Academic Publishers, 1991, p. 229-271.

[50] H. Gharavi, A. Tabatabai: *Subband coding of digital images using two-dimensional quadrature mirror filtering.* Proc. SPIE, 1986.

[51] H. Gharavi, A. Tabatabai: *Subband coding of monochrome and color images.* "IEEE Trans. on Circuits and Systems", Vol. 35, February 1988, p. 207-214.

[52] K. N. Ngan, W. Chooi: *Very low bit rate video coding using 3d subband approach.* "IEEE Trans. on Circuits and Systems for Video Technology", Vol. 4, June 1994, p. 309-316.

[53] C. Podilchuk, N. Jayant, N. Farvardin: *Three dimensional subband coding of video.* "IEEE Trans. on Image Processing", Vol. 4, February 1995, p. 125-139.

[54] J. Katto, J. Ohki, S. Nogaki, M. Ohta: *A wavelet codec with overlapped motion compensation for very low bit-rate environment.* "IEEE Trans. on Video Technology", Vol. 4, June 1994, p. 328-338.

[55] A. Gersho, R. Gray: *Vector quantization and signal compression.* Kluwer Academic Publishers, 1992.

[56] L. Torres, J. Huguet: *An improvement on codebook search for vector quantisation.* "IEEE Transactions on Communications", Vol. 42, February 1994, p. 208-210.

[57] P. Strobach: *Tree-structured scene adaptive coder.* "IEEE Transactions on Communications", Vol. 38, April 1990, p. 477-486.

[58] J. Vaisey, A. Gersho: *Image compression with variable block size segmentation.* "IEEE Tr. on Signal Processing", Vol. 40, August 1992, p. 2040-2060.

[59] E. Shustermann, M. Feder: *Image compression via improved quadtree decomposition algorithms.* "IEEE Transactions on Image Processing", Vol. 3, March 1994, p. 207-215.

[60] R. J. Clarke: *Transform coding of images.* Academic Press, 1985.

[61] A. Palau, G. Mirchandani: *Image coding with discrete cosine transforms using efficient energy-based adaptive zonal filtering.* "IEEE ICASSP", 1994, p. V337-V340.

[62] H. Yamaguchi: *Adaptive dct coding of video signals.* "IEEE Transactions on Communications", Vol. 41, October 1993, p. 1534-1543.

[63] K. Ngan: *Adaptive transform coding of video signals.* "IEE Proceedings", Vol. 129, February 1982, p. 28-40.

[64] R. Clarke: *Hybrid intra-frame transform coding of image data.* "IEE Proceedings", Vol. 131, February 1984, p. 2-6.

[65] F.-M. Wang, S. Liu: *Hybrid video coding for low bit-rate applications.* "IEEE ICASSP", 1994, p. 481-484.

[66] M. Ghanbari, J. Azari: *Effect of bit rate variation of the base layer on the performance of two-layer video codecs.* "IEEE Transactions on Communications for Video Technology", Vol. 4, February 1994, p. 8-17.

[67] R. A. Salami, L. Hanzo, R. Steele, K. H. J. Wong, I. Wassell: *Mobile radio communications, ch. Speech Coding.* London: IEEE Press-Pentech Press, 1992, p. 86-346.

[68] M. Narasimha, A. Peterson: *On the computation of the discrete cosine transform.* "IEEE Transactions on Communications", Vol. 26, June 1978, p. 934-936.

[69] L. Torres: *Segmetation based coding of textures using stochastic vector quantization.* "IEEE ICASSP", 1994, p. 5-597 - 5-600.

[70] M. Y. Jaisimha, J. R. Goldschneider, E. Alexander: *On vector quantization for fast facet edge detection.* "IEEE ICASSP", 1994, p. 5-37 - 5-40.

[71] P. Yu, A. Venetsanopoulos: *Hierarchical finite-state vector quantisation for image coding.* "IEEE Transactions on Communications", Vol. 42, November 1994, p. 3020-3026.

[72] C.-H. Hsich, K.-C. Chuang, J.-S. Shue: *Image compression using finite-state vector quantization with derailment compensation.* "IEEE Transactions on Circuits and systems for Video Technology", Vol. 3, October 1993, p. 341-349.

[73] N. Nasrabadi, C. Choo, Y. Feng: *Dynamic finite-state vector quantisation of digital images.* "IEEE Transactions on Communications", Vol. 42, May 1994, p. 2145-2154.

[74] V. Sitaram, C. Huang, P. Israelsen: *Efficient codebooks for vector quantisation image compression with an adaptive tree search algorithm.* "IEEE Transactions on Communications", Vol. 42, November 1994, p. 3027-3033.

[75] L. Lu, W. Pearlman: *Multi-rate video coding using pruned tree-structured vector quantization.* "IEEE ICASSP", 1993, p. V-253-V256.

[76] W. Yip, S. Gupta, A. Gersho: *Enhanced multistage vector quantisation by joint codebook design.* "IEEE Transactions on Communications", Vol. 40, November 1992, p. 1693-1697.

[77] L. Po, C. Chan: *Adaptive dimensionality reduction techniques for tree-structured vector quantisation.* "IEEE Transactions of Communications", Vol. 42, June 1994, p. 2246-2257.

[78] F. Bellifemine, R. Picco: *Video signal coding with dct and vector quantisation.* "IEEE Transactions on Communications", Vol. 42, February 1994, p. 200-207.

[79] K. Ngan, K. Sin: *Hdtv coding using hybrid mrvq / dct.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 3, August 1993, p. 320-323.

[80] D. Kim, S. Lee: *Image vector quantiser based on a classification in the dct domain.* "IEEE Transactions on Communications", April 1991, p. 549-556.

[81] B. Ramamurthi, A. Gersho: *Classified vector quantization of images.* "IEEE Transactions on Communications", Vol. COM-34, November 1986, p. 1105-1115.

[82] X. Zhang, M. Cavenor, J. Arnold: *Adaptive quadtree coding of motion-compensated image sequences for use on the broadband isdn.* "IEEE Transactions on Circuits and Systems for Video Technology", Vol. 3, June 1993, p. 222-229.

[83] P. Cherriman, L. Hanzo: *Robust h.263 video transmission over mobile channels in interference-limited environments.* Proc of First International Workshop on Wireless Image/Video Communications, (Loughborough, UK), 4-5 September 1996, p. 1-7.

[84] P. Cherriman, L. Hanzo: *Power-controlled h.263-based wireless videophone performance in interference-limited scenarios.* PIMRC'96, Taipei, Taiwan, 15-18 October 1996, p. 158-162.

[85] P. Cherriman, L. Hanzo: *Programmable h.263-based wireless video transceivers for interference-limited environments.* Proc. of IEEE ICCS'96 / ISPACS'96, Vol. 3, (Singapore), 25-29