Fixed-rate Video Codecs for Wireless Systems *†

J. Streit, L. Hanzo
Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
Tel: +44 1703 593 125, Fax: +44 1703 593 045
Email: jss@ecs.soton.ac.uk, lh@ecs.soton.ac.uk
http://www-mobile.ecs.soton.ac.uk

July 19, 1996

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. These 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 kbps, respectively.

1 Motivation

The theory and practice of image compression has been consolidated in a number of established monographs, such as for example Reference [1] by Jain. A plethora of video codecs have been proposed in the excellent special issues edited by Tzou, Mussmann and Aizawa [2] as well as Girod et al [3] for a range of bitrates and applications, but the individual contributions by a number of renowned authors are too numerous to review. As an attractive 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 [4, 5, 6, 7, 8, 9].

The outline of the paper is as follows. Section 2 describes the gain-cost quantised, fixed but arbitrarily programmable rate discrete cosine transformed (DCT) video codec, while Sections 3 and 4 highlight the vector-quantised (VQ) and quad-tree (QT) coded schemes. Our conclusions are presented in Section 5.

2 Low Bitrate DCT Codecs

2.1 DCT Codec Schematic

The proposed programmable codec was designed to switch between intra- and inter-frame modes of operation, as seen in Figure 1. At the commencement of communications we transmit a low-resolution intra-frame coded frame, in order to initialise 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. Specific algorithmic details of the codec were documented in Reference [4], hence here we refrain from elaborating on the motion active/passive and DCT active/passive block classification technique employed, which operates under the instructions of the bitrate control algorithm of Figure 1. We found that earmarking about 30-40 of the 396 8×8 QCIF blocks as motion-active and 30-40 as DCT-active was a good compromise in terms of video quality and implementational complexity for bitrates around 10 kbps. The side-information represented by the corresponding activity-tables was amenable to further compression

*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

†1st IEEE Wireless Video Communications Workshop, 4-5 Sept. 1996, Loughborough, UK

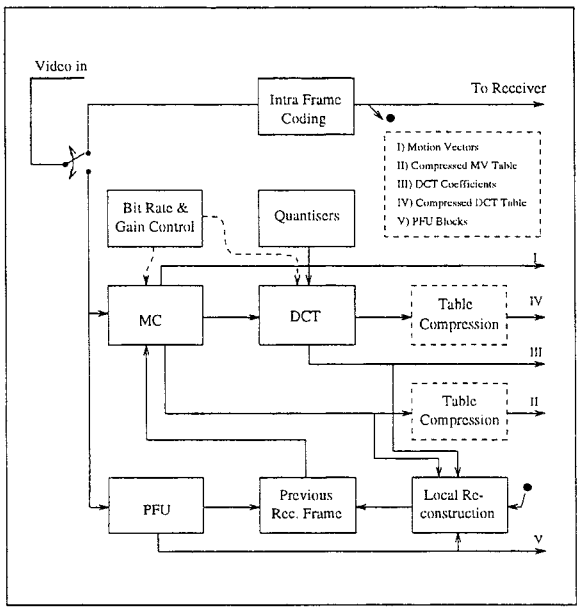

Figure 1: Schematic of the Multi-Class DCT codec

at the cost of reduced robustness against channel errors, which is indicated by the optional activity-table compression printed in broken lines in Figure 1.

2.2 Adaptive Bit Allocation Strategy

The adaptive codec's bitallocation is summarised in Table 1. We found that the best subjective and objective videophone quality was achieved, when the number of active blocks for the motion compensation (MC) and DCT was roughly the same, although not necessarily the same 30 blocks were processed by the two independent algorithms. For the limited search scope of head-and-shoulders videophone sequences the encoding of the MVs requires only 4 bits per active block, while that of the DCT coefficients needs 12 bits/block, including the quadruple-class 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 for the Partial Forced Up-date (PFU) procedure employed to improve the codec's robustness, as detailed in Reference [4] . The PFU was typically configured to refresh 22 out of the 396 blocks in each frame. Therefore $4 \times 22 = 88$ 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 kbps 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 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 Figure 2 reveals, DCTC1 at 8 kbps achieves a similar quality to that of DCTC2 at 11.3 kbps.

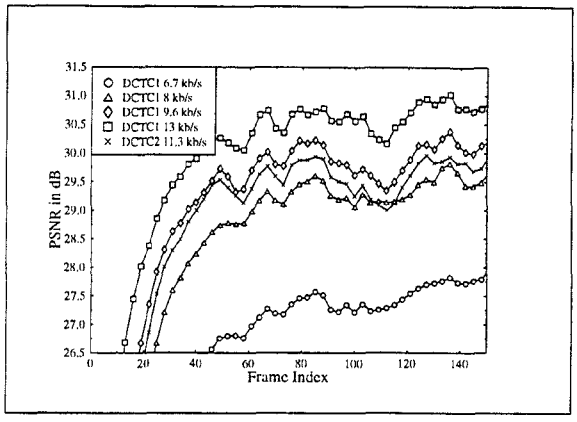

Figure 2: PSNR versus frame index performance of DCTC1 at various bit rates and for DCTC2 at 11.3 kb/s for the 'Claire' sequence

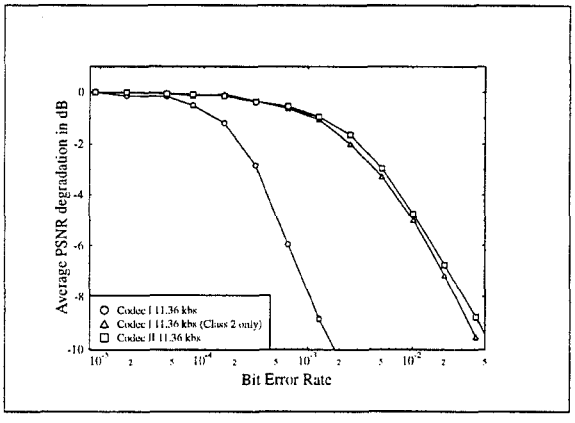

Figure 3: PSNR degradation versus BER for DCTC1 and DCTC2

2.3 DCT Codec Robustness

The performance of DCTC1 was tested at 10 frames/s and 6.7, 8, 9.6 and 13 kbps, which are the speech rates of the Japanese Digital Cellular [10], American IS-54 [11], IS-95 [12] and the Pan-European GSM system [13], respectively. The results for DCTC2 are similar at a 35 % higher bit rate. [1]

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 Figure 3. 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 BERs up to $2 \cdot 10^{-2}$. In reference [4] we proposed an appropriate transmission scheme, which takes advantage of this characteristic. As evidenced by Figure 3, 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 reference [4]. Having studied the algorithmic and performance

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

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

3 Vector-quantised Video Codecs

Vector quantisation (VQ) is a generalisation of scalar quantisation, a technique lavishly documented in an excellent monograph by Gray and Gersho [14]. Algorithmic details of the proposed VQ scheme were detailed in Reference [6], hence here we concentrate mainly on the performance of the codec. The VQ codec's schematic is akin to that of the DCT codec shown in Figure 1, with the exception of invoking a specially trained codebook for representing the 8×8 pixel block instead of encoding their DCT coefficients.

Similarly to the previously proposed DCT-based codecs, we contrived two VQ schemes, 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 so-called classified VQ principle, using a codebook size of 256 which lead to an overall codec complexity of around 15 Mflops, when employing the previously described active / passive block classification. The MCER was generated for all 396 8×8 blocks and a bit-rate constrained fraction of the highest-energy 20-50 % MCER blocks were vector quantised.

The peak signal-to-noise ratio (PSNR) versus frame index performance of the VQC1 scheme is portrayed in Figure 4 for the 'Claire' sequence at the previously introduced 2nd generation mobile radio speech bit rates of 6.7, 8, 9.6 and 13 kbps. Lastly, the associated bit allocation schemes are summarised in Table 1 in contrast to our other prototype codecs. [2] This Table reveals a range of interesting aspects, showing for example that while the DCT codec allocated 12 bits/block for DCT-based MCER coding, for a similar quality the VQ scheme required only an 8-bit, 256-entry codebook. 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 [6] is akin to that of DCTC1 and DCTC2, respectively, which was shown in Figure 3. These issues will be comparatively studied for DCT, VQ and QT codecs in our Conclusions Section. Let us now briefly consider quad-tree (QT) coded schemes.

Codec	FAW	PFU	MV Index + MV	DCT Ind. + DCT	VQ Ind. + VQ	QT + PC	Padding	Total
DCTC1	22	22×4	30×9 + 30×4	30× 9 + 30×12	-	-	6	1136
DCTC2	22	22×4	< 350 (VLC)	< 350 (VLC)	-	-	VLC	800
VQC1	22	22×4	38×9 + 38×4	-	31×9 + 31×8	-	5	1136
VQC2	22	22×4	< 350 (VLC)	-	< 350 (VLC)	-	VLC	800
QTC1	22	20×4	< 500 (VLC)	-	-	< 565 + 1 or 80	VLC	1136

Table 1: Bit Allocation Table

4 Quad-tree based Codecs

The proposed QT codecs also obey the structure of Figure 1, but the DCT-based MCER compression was replaced by QT-based compression. Again, for reasons of space economy, here we refrain from detailing the algorithmic design of the fixed-rate cost-gain quantised QT codec, the interested reader is referred to [5] for a detailed discussion. The codec's bit allocation scheme is summarised in Table 1 in contrast to our other benchmarkers.

Suffice to say here that upon assessing the potential of a number of different approaches to contriving an appropriate adaptive bit allocation scheme we finally arrived at Algorithm 1 [5]. 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'. 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

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

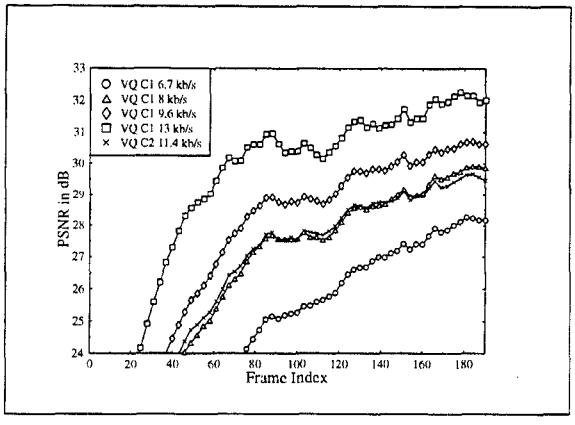

Figure 4: PSNR versus frame index performance of VQC1 at various bit rates and for VQC2 at 11.3 kb/s for the 'Claire' sequence

Algorithm 1 *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 [5].*

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.

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.

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 bitrate performance of the QT-codec will be shown in Figure 6 in contrast to the DCT- and VQ-based schemes [3], while its robustness evaluated in terms of PSNR versus BER will be compared to that of our other benchmarkers in Figure 8. Having highlighted the salient features of the proposed QT codec let us now focus our attention on the performance comparison of the schemes considered.

5 Discussion and Conclusions

We comparatively studied five different fixed-rate QCIF video codecs suitable for wireless videotelephony and studied their robustness. The corresponding bitallocation schemes were summarised in Table 1. The associated transmission issues, including source sensitivity-matched forward error correction (FEC) coding, adaptive modulation and automatic repeat request (ARQ) schemes have been discussed in depth in a series of companion papers published by the authors in references [4], [6], [5] and [8].

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

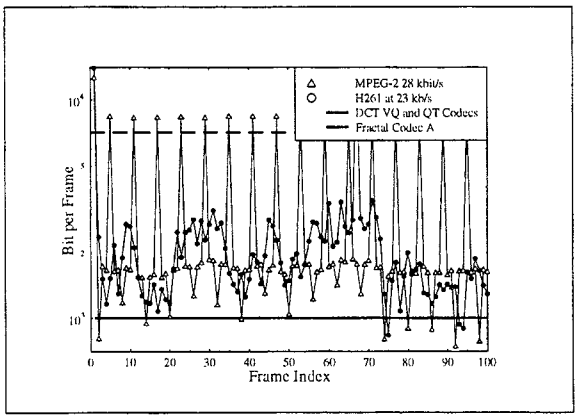

Figure 5: Bit rate fluctuation versus frame index for the proposed adaptive codecs and two standard codecs

Let us finally compare our proposed inter-frame codecs to two widely used standard codecs, namely the MPEG-2 and H261 codecs. The latter standard schemes are typically variable rate codecs, which make extensive use of variable-length compression techniques, such as RL-coding and entropy coding [1], although it is possible to invoke appropriate adaptive packetisation and multiple encoding operations in order to arrive at a required near-constant bitrate. 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. The voice signal's latency or delay is becoming annoying for delays of 100 ms, hence perfect lip-synchronisation cannot be realistically achieved.

The above standard codecs also require the transmission of at least one intra-frame coded frame at the commencement of transmission in order to provide a reference for the operation of the motion compensation. The transmission of 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 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 forced update scheme was more appropriate for the targeted mobile radio applications.

In our experiments portrayed in Figure 5 we stipulated a fixed bit rate of 10 kbps 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. The corresponding PSNR curves are displayed in Figure 6. Observe in Figure 5 that the number of bits / frame for our proposed codecs is always 1000, corresponding to 10 kbps 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.

A direct comparison of the above five codecs in Figures 5, 6 and 8 reveals the following findings:

- Our codecs achieve a similar performance to the MPEG-2 codec at less than half the bit rate. The H-261 codec at 22 kb/s, ie more than twice the bit rate, outperforms our codecs by about 2 dB in terms of PSNR. Note furthermore that our fixed-rate DCT and VQ codecs require about 20 frames to reach their steady-state video quality due to the fixed bit rate limitation, which is slightly prolonged for the QT codec.

- 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.236 and MPEG-2 codecs may stretch to several frames due to the P-frames. In order

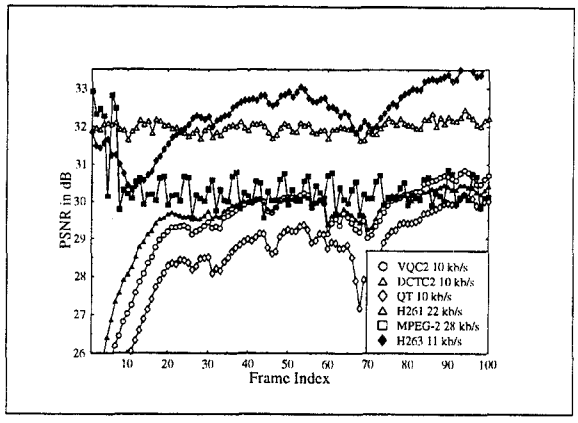

Figure 6: PSNR versus frame index performance of the proposed adaptive codecs and three standard codecs

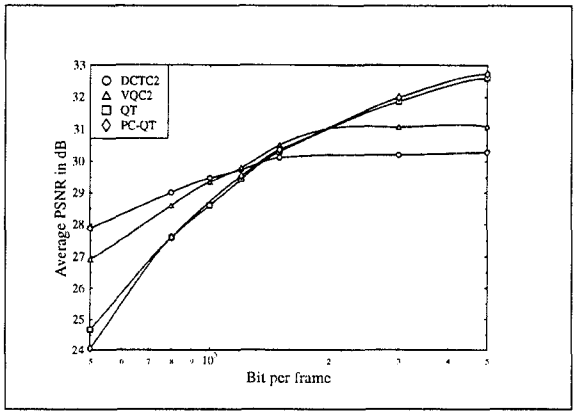

Figure 7: PSNR versus bit rate performance of the proposed adaptive codecs

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 runlength-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.

6 Acknowledgement

The financial support of the EPSRC, UK in the framework of the research contract GR/K 74043 is gratefully acknowledged.

References

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

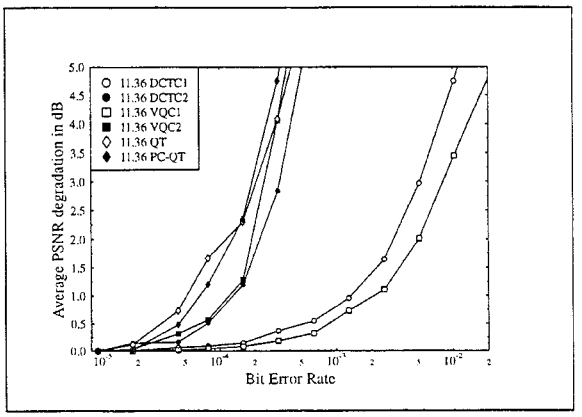

Figure 8: PSNR degradation versus BER for the proposed codecs

[2] Guest Editors: K. Tzou, H. Mussmann, and K. Aizawa, "Special issue on very low bit rate video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, pp. 213–357, June 1994.

[3] Guest Editors: B. Girod et al, "Special issue on image sequence compression," *IEEE Transactions on Image Compression*, vol. 3, pp. 465–716, September 1994.

[4] L. Hanzo and J. Streit, "Adaptive low-rate wireless videophone schemes," *IEEE Transactions Video Technology*, August 1995.

[5] J. Streit and L. Hanzo, "Quad-tree based parametric wireless videophone systems." submitted to IEEE Transactions Video Technology.

[6] J. Streit and L. Hanzo, "Vector-quantised low-rate cordless videophone systems." submitted to IEEE Transactions Vehicular Technology.

[7] R. Stedman, R. Steele, and L. Hanzo, *Multi-media Technologies and Future Applications*, ch. A portable multi-media communicator scheme, pp. 31–54. Pentech Press, 1994.

[8] J.Streit and L.Hanzo, "A fractal video communicator," in *IEEE Conference on Vehicular Technology*, (Stockholm, Sweden), pp. 1030–1034, June 1994.

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

[10] Research and Development Centre for Radio Systems, Japan, *Public Digital Cellular (PDC) Standard, RCR STD-27.*

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

[12] 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.

[13] L.Hanzo and J. Stefanov, *Mobile Radio Communications*, ch. The Pan-European Digital Cellular Mobile Radio System - known as GSM, pp. 677–773. Pentech Press, London, 1994.

[14] A. Gersho and R. Gray, *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, 1992.