Peer reviewed version

Link to published version (if available):
10.1109/ICIP.2001.958670

Link to publication record in Explore Bristol Research
PDF-document

ROBUST AND SCALABLE VIDEO COMPRESSION USING MATCHING PURSUITS AND ABSOLUTE VALUE CODING

David Redmill

Centre for Communications Research, University of Bristol, BS8 1UB, U.K.

ABSTRACT

The main objective of this paper is to develop a fine grained scalable video codec which can fully exploit temporal correlation's within all layers, and thus offer performance closer to that of a single layer non-scalable system. To this end we propose the use of a combination of matching pursuits together with absolute value coding of the displaced frame difference information.

The proposed codec is also shown to be highly robust to channel errors. This results from a combination of the use of fixed length code-words and a reduction of temporal error propagation. The use of fixed length code-words also facilitates the use of lower complexity transcoding operations to support more sophisticated forms of scalability.

1. INTRODUCTION

Recent years have seen a increase of the number of available communications and networking systems operating with a wide range of band-width and error conditions. However, many standard video codecs (e.g. H..263, MPEG-1) are designed to operate at a single bit-rate without any transmission errors. More recently codecs supporting a number of coding layers have emerged, including MPEG-2, H.263+ and MPEG-4. These have the advantage that the same bit-stream can be used at more than one (typically 2 or 3) bit-rates. However, this is typically achieved at the expense of the compression performance for the higher rate layers. A further development is the introduction of fine-grained scalable (FGS) systems (e.g. MPEG-4), enabling a single bit-stream to be decoded across a nearly continuous range of bit-rates. One problem with many existing FGS systems is the use of intra-frame mode for the enhancement layer, which limits the compression achievable as a result of not fully exploiting temporal correlations.

The problem encountered is that if motion compensated prediction were used within the enhancement layer, then the truncation of the bit-stream during the scaling process can lead to errors which propagate temporally to subsequent frames. This severely limits system performance. The same temporal error propagation problem is also encountered when designing robust encoders, where it is the effects of channel errors can propagate temporally.

Temporal error propagation is typically limited by the use of intra-frame coded portions or complete frames. The disadvantage of intra-frame coding is the overhead incurred, compared to inter-frame coding. Another alternative is the use of a loop filter to attenuate high frequency errors without excessive coding overhead. A similar effect can also be achieved by using an overlapped motion compensation technique, which can also improve the prediction quality and thus reduce the bit rate. In this paper we propose the use of a new technique referred to as absolute value coding [1].

2. MATCHING PURSUITS (MP)

An important component for any motion compensated video codec is the coding of the displaced frame difference (DFD) or prediction error. Many typical systems use either a DCT or wavelet transform in conjunction with zig-zag scanning, run-length and entropy coding. A problem with this approach is that it cannot easily be used for fine-grained scalable systems.

A more recent alternative is the use of embedded zero-tree coding or the SPIHT algorithm [2] in conjunction with a wavelet transform. While these systems work well for still images, they are less suitable to DFD images due to the poorer statistics. DFD signals tend to be dominated by few fine edges.

More recently a new technique has been proposed known as matching pursuits (MP) [3,4,5]. In MP the DFD signal is recursively approximated by a series of atoms, comprising scaled and shifted basis functions, taken from a pre-determined dictionary. The recursive approximation technique of MP is well suited to scalable systems. One problem with MP is the relatively high complexity incurred in the encoder.

In this paper we use a block-based MP system. The dictionary consists of separable basis functions, which are restricted to individual blocks (typically 16x16 for luminance or 8x8 for chrominance). The MP search is

performed by first finding the best atom for each block, and then searching for the best overall. For subsequent atoms only the block corresponding to the previously coded atom needs to be re-searched. This approach significantly reduces the complexity to an acceptable level. It also enables the use of intra-frame coding of selected blocks.

3. ABSOLUTE VALUE CODING (AVC)

The fundamental idea behind absolute value coding (AVC) [1] is that the absolute value for each atom (or coefficient) is coded rather than a differential value. This effectively eliminates any temporal error propagation to the coded atoms (or coefficients). Thus whenever an atom/coefficient is received at the decoder it can be accurately decoded irrespective of any possible incorrect prediction. The atom/coefficient selection procedure at the encoder is still based on the DFD signal. The use of AVC significantly reduces temporal error propagation to active regions of the image. However, significant temporal error propagation can still occur in stationary background regions. This is subjectively much less objectionable. The disadvantage of AVC is an increase in the number of bits required to code these magnitude values (since the absolute value of an atom has a larger variance than the differential value). Although AVC can be used with a variety of coding systems including the DCT, it is well suited for use with matching pursuits where the ratio between positional and magnitude information is relatively high, and consequently the overhead incurred by AVC is less significant.

5. PROPOSED CODEC

The proposed codec uses a 16x16-pixel macro-block (MB) structure. Firstly the intra-inter mode selection and motion vectors are coded using 1 bit/MB for mode selection and the H.263 Huffman codes for the motion vectors. The DFD signal is then coded as a series of atoms coded as fixed length code-words. Each atom comprises a MB number, a horizontal and vertical dictionary index and a quantization index. For the QCIF sequences tested, 31 bits per code-word are used. The use of fixed length coding has several advantages including robustness to channel errors, and ease of intermediate processing. This facilitates the use of low complexity transcoding systems to select which atoms to transmit (e.g. based on region of interest, spatial frequency or chrominance/luminance).

In order to provide good performance across a range of scaled bit-rates, the encoder uses a layered structure and selects atoms according to one of several DFD signals achieved by decoding at different rates

(25,50,100 & 200 Kb/s for the results shown). Note that the decoder only uses a single layer (with the exception of the enhancement system, which uses a separate base layer).

6. RESULTS

Figure 1 shows PSNR against bit-rate curves for five different systems using the Forman sequence coded at QCIF resolution and 12.5 frames/s. The top-most curve shows results is for H.263 (non-scalable). The next curve shows results for the proposed system using a single layer codec (i.e. non-scalable). This curve exhibits inferior performance compared to H.263. This is largely due to use of simple fixed length coding rather than more sophisticated entropy coding. Results could also be improved by more careful selection of the matching pursuits dictionary.

The third and fourth curves (in figure 1) show scalable results achieved for the proposed fine grained scalable system, both with and without AVC. Both curves exhibit their best performance at the bit-rates used by the encoder (25,50,100 & 200 Kb/s). In between these rates, the performance dips as a result of temporal error propagation. The dips for the system using AVC are much less pronounced as a result of the reduction in temporal error propagation. Another interesting point is that successive peaks are higher for the system using AVC than without. This is a result of not having to code similar information in several layers. Thus AVC can improve the performance of a layered scalable system. A final point to note is that the performance for the fine grained scalable system suffers from the use of a single motion field for the whole range of bit-rates. Thus the motion compensation is generally sub-optimal compared to a single rate system.

The final curve (in figure 1) shows a more traditional fine-grained scalable system using a base layer and an intra-coded (predicted from base layer) enhancement layer. This curve exhibits significantly lower performance at higher rates. This is a result of using a poor quality temporal prediction (from the base layer). Thus a lot of high frequency information needs to be re-coded for each frame.

In order to determine the potential improvements that might be made using a more sophisticated entropy coding system, we have utilized the error resilient positional code (ERPC) [5,6] to improve the positional coding of the atoms. This is still a fixed length coding system and thus still robust and simple to decode. Figure 2 shows results using the ERPC. The graph shows an improvement compared to figure 1 of about 1dB for higher rates. One disadvantage of using the ERPC is that it removes the code-word ordering required for effective scalability. Thus if the ERPC is used in a fine grained

1022

scalable system, it should be used in the scaling process rather than the encoder.

The use of fixed length code-words makes the proposed DFD codec very robust to channel errors. Figure 3 shows results obtained with a random bit error rate of 0.1% for all the DFD data. The results for both the matching pursuits systems are superior to those for H.263. This is primarily due to the use of fixed length coding. The codec using AVC suffers from significantly less error propagation as expected. The motion vector information has not been corrupted since currently it is not robustly coded. A practical system would need to use either channel coding or an alternative coding method for the motion information.

Figure 4 shows the 25th frame (2 seconds) after a lost intra frame. The figure demonstrates how without using AVC, the missing low frequency information for the first frame can propagate indefinitely. However, the proposed system using AVC can quickly recover from even the most severe error. Errors can still be seen in the slow moving background and hat regions. However these are subjectively much less objectionable.

6. CONCLUSIONS

In this paper we have proposed a fine-grained scalable video codec using matching pursuits and absolute value coding (AVC). The proposed codec utilizes motion compensated prediction across the whole range of bit-rates, rather than just the base layer. This is made possible due to the reduction in temporal error propagation offered by AVC. Results are shown which demonstrate how the proposed system can achieve a performance much closer to that of a single layer system. The proposed system utilizes a fixed length coding strategy. This makes it both highly robust to channel errors and facilitates simple transcoding to support other forms of scalability such as region of interest. Results are also presented to demonstrate how the proposed system exhibits a high resilience to channel errors and quick recovery from even the most severe errors, such as the loss of an intra-frame.

The proposed system is still the subject of ongoing research and development. In particular improvements might be made in the choice of dictionary and the use of more sophisticated entropy coding. The improvement of the motion vector coding also requires attention in order to improve its robustness to channel errors.

7. REFERENCES

[1] D. W. Redmill and D. R. Bull, "Resilient video compression using absolute value coding", In proc IEE International Conference on Image Processing and its Applications, pp. 591-595, Machester UK, July 1999.

[2] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees", IEEE Transactions on Circuits and Systems for Video Technology, 6, pp. 243-249, 1996.

[3] R. Neff and A. Zakor, "Very low bit-rate video coding based on matching pursuits", IEEE Transactions on Circuits and Systems for Video Technology, 7, pp. 158-171, 1997.

[4] P. Czerepiński, C. Davies, C. N. Canagarajah and D. R. Bull, "Matching pursuits vide coding: Dictionaries and fast implementation" IEEE Transactions on Circuits and Systems, 26, 1pp 1103-1115, October 2000.

[5] M. F. Tariq, P. Czerepiński, A. Nix, D. Bull and C. N. Canagarajah, "Robust and scalable matching pursuits video transmission using the Bluetooth air interface standard", IEEE Transactions on Consumer Electronics, 46, pp. 673-681, August 2000.

[6] N. T. Cheng and N. G. Kingsbury, "The ERPC: An efficient error-resilient technique for encoding positional information of sparse data", IEEE Transactions on Communications, 40, pp. 140-148, 1992.
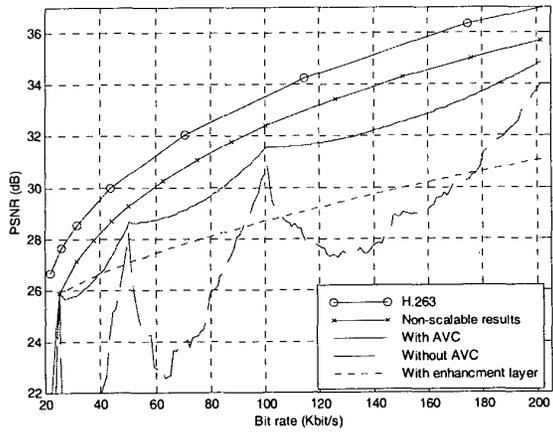
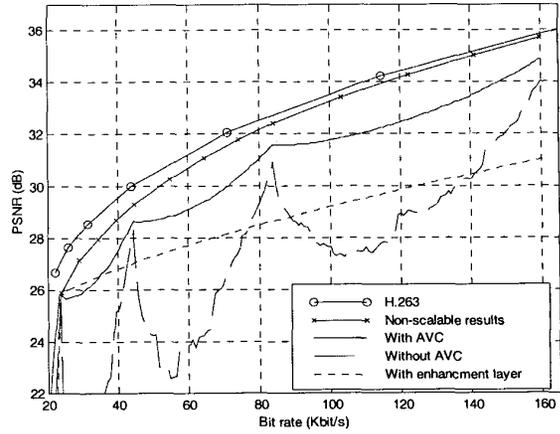Figure 1: Performance curves without ERPC.

Figure 2: Performance cureves with ERPC.

Figure 3: Effects of 0.1% BER on the DFD coded data. Frame 50 a) H.263, b) MP without AVC, c) MP with AVC.

Figure 4: Recovery from a lost intra-frame. a) H.263, b) MP without AVC, c) MP with AVC.

1024