A HYBRID VIDEO CODER BASED ON H.264 WITH MATCHING PURSUITS

Haoxiang Zhang, Xiaopeng Wang, Wei Huo, and D. M. Monro
Department of Electronic and Electrical Engineering,
University of Bath, Bath, United Kingdom,
{ eephz, eepxw, eepwh, d.m.monro } @bath.ac.uk

ABSTRACT

We present a novel video coding system based on the motion compensation and estimation model of H.264 and a hybrid wavelet/matching pursuits image coder for residual frames. After wavelet pre-transformation, a high performance matching pursuits codebook is used to find a series of atoms which are coded by the MERGE coder. The combination of these two techniques produces a highly embedded data stream that can be easily truncated at any desired point. Optimum settings of some key parameters are investigated and a series of experiments are performed. The system works very efficiently and can outperform H.264 at low and medium bit rates.

1. INTRODUCTION

This paper presents a hybrid video coder based on H.264 incorporating a Matching Pursuits (MP) residual image coder from the Bath University Matching Pursuits (BUMP) project. In most hybrid video coders, a motion estimation residual image, or Displaced Frame Difference (DFD), is coded by methods similar to those for coding intra frames. However DFD images usually consist of localized sparse islands of data over a low level noisy background. Neff and Zakhor demonstrated that MP can code DFD images efficiently by achieving improved low bit rate video coding by MP for motion compensated residual images within an H.263 video codec [1].

In the BUMP project, [2-4], large gains in fidelity accompanied by significant reductions in complexity were achieved in MP for coding both still images and motion compensated residuals. These advances came from pre-transformation by wavelet [2], an embedded coding scheme for MP, and improved dictionaries found by 'basis picking' [3]. A result of particular importance was that small dictionaries of narrow bases performed well in the hybrid Wavelet/MP codec [4].

2. WAVELET/MP DECOMPOSITION

The wavelet/MP codec developed in BUMP is used for DFD image coding in this study. For coding, a multiscale wavelet decomposition is applied using the well known biorthogonal 9/7 filterbank before MP approximation. It has been shown that 2 scales for CIF (352 x 288) residuals and 5 scales for D1 (704 x 576) still images are a good choice [3]. Atoms for MP are then found directly on the 2D wavelet coefficient array using a dictionary of 2D bases.

In MP coding a dictionary of basis functions is repeatedly searched for the inner product of largest magnitude within the data set. The position, sign, quantized amplitude and dictionary index of the selected basis form the code of an 'atom'. The atom is subtracted from the data before the process is repeated on the remaining residual. Often the bases are an over-complete, non-orthogonal set of Gabor functions, and the importance of the codebook in the performance of the algorithm is well established [3, 4].

Once the wavelet transform has been done, the MP algorithm can be described as a three-step per atom process:

Initialise compute a full set of inner products between the wavelet coefficients and all bases in a codebook.

Repeat

1. *Find atom*. Full search or reduced complexity strategy.
2. *Image Update*. Subtract atom from data.
3. *Repair Inner Products*. Recompute required inner products only in atom footprint.

Until distortion or bit rate criterion met

3. THE EMBEDDED PLQ/MERGE CODER

Precision Limited Quantization (PLQ) [6] fits well with embedded coding schemes, and has been found useful in maximizing PSNR in MERGE coding both with audio and video data [2].

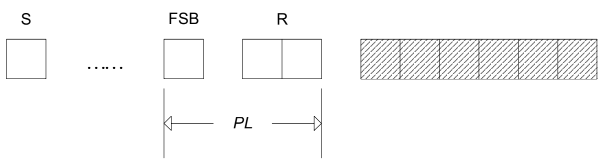

Figure 1. Precision Limited Quantization of an atom of amplitude A by the triple $\langle S,F,R\rangle$.

In Figure 1, if A is the unquantized amplitude of an atom, then $S = sign(A)$, $F = \log_2|A|$, i.e. the First Significant Bit (FSB) of $|A|$, and $R =$ the remaining bits in the range 0 to $2^{PL-1}-1$, for precision limit $PL > 1$, including the FSB.

Each atom chosen by MP has a location in the space of wavelet coefficients with the *S*, *F* and *R* attributes and codebook index *K*. Lossless coding of the attributes is done by the MERGE algorithm, in which atoms are gathered into groups with attributes in common, and the positions are signalled by run length coding. This works well with PLQ=2, which keeps the number of groups reasonably small. To reduce the number of groups further, the sign *S* of each atom is sent as one bit of side information which is efficient because positive and negative signs have near-equal probability.

If the Precision Limit is *PL*, the MERGE algorithm is:

For *F* (the FSB) from Maximum to Minimum
  For *R* (the amplitude Residual) from 2^{PL-1} to 0
    For each Basis Function *K* used
      Signal by Run Length Coding the position of each atom with attributes (*F, R, K*).
      Send the Sign *S* of the atom (1 bit)
    End of *K* (Basis Function) Group
  End of *R* (PLQ Residual) Group
End of *F* (FSB) Group

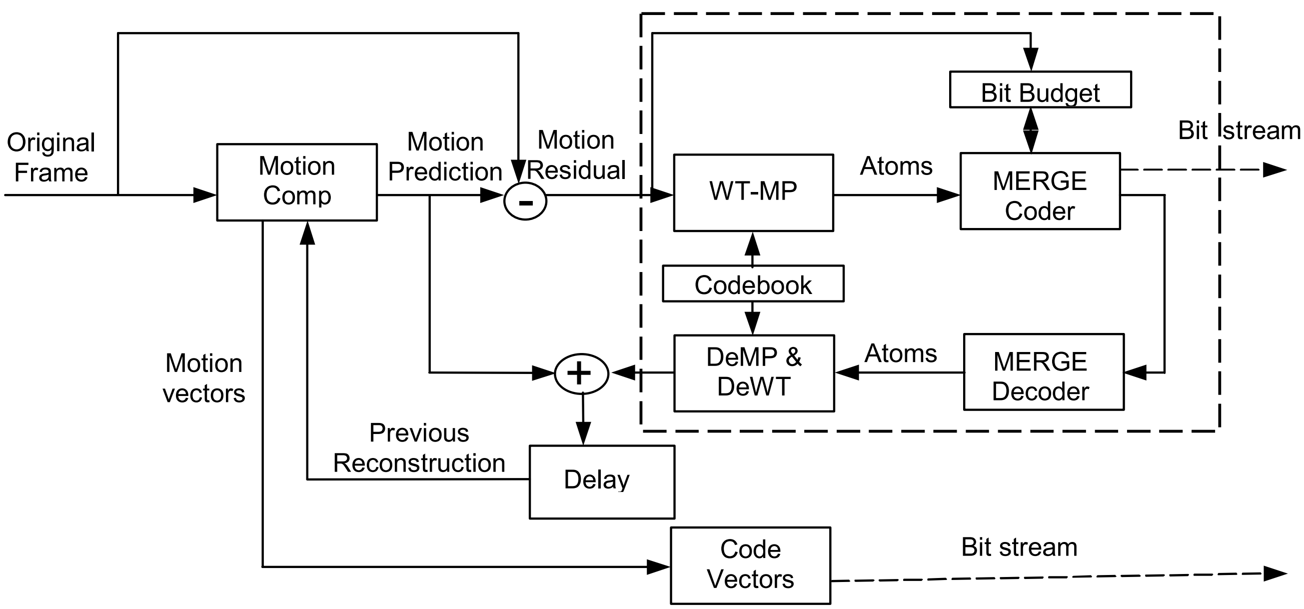

Figure 2. Structure of the proposed video coding system.

Maximum embedding is achieved by sending atoms in order of decreasing amplitude, with the codebook entry as the innermost loop. MERGE automatically compensates for variations in the frequency of occurrence of the attributes of the atoms and eliminates the need for entropy coding of them. The adaptive run length coding in MERGE adjusts to the statistics of the atom position.

4. THE HYBRID VIDEO CODER

The present video coder utilizes the publicly available H.264 reference code [5] as a framework, which has an advanced motion estimation module. We use this module to produce both motion vectors and DFD frames. In H.264, for each macroblock in a DFD frame, there are several candidate coding modes {Intra, 16x16, 8x16, 16x8, 8x8, 4x8, 8x4, and 4x4}. For efficient Intra block coding, H.264 uses a prediction scheme, in which the pixels in a block are predicted from the available reconstructed pixels in adjacent blocks. However in an MP coder, each residual frame is treated as a single unit, so that prediction between blocks is not possible. To resolve this difference the H.264 Intra mode is disabled for DFD frames in these experiments. Apart from this, most of the advantages of the H.264 motion module are preserved, including the variable block sizes, quarter pixel motion vector accuracy, and usage of multiple reference frames.

The structure of the video coder is showed in Figure 2, with the BUMP DFD encoder specified in the dashed rectangle with PL=2. As can be seen in Figure 2, the original frame is compared to the previously reconstructed frames to produce motion vectors and the DFD frame. The latter is taken as input to the BUMP coder, and the reconstruction is passed back for future motion estimation.

For comparison, we enable the rate control mechanism in the H.264 code, and set it at frame level for CIF. The number of bits allocated by H.264 for each frame is recorded, so the bit budget for the proposed BUMP coder can be calculated by subtracting the number of unavailable bits from this recorded number:

Budget = Recorded_bits – Unavailable_bits

where the *Unavailable bits* includes those for coding motion vectors, block modes, etc.

The DFD coder first finds a fixed number of atoms, MaxAtoms, which would normally be in excess of the expected requirement for the desired bit budget. The MERGE coder can then code an embedded data stream from those atoms to match the target bit rate. Although the bit stream created by the MERGE coder at each iteration would be very close to the desired size, it is not exact as the finest unit it can achieve is the cost of one atom. There are usually a few bits to spare, so for accurate rate control the actual number of bits is passed back to the Bit Budget controller so that any spare bits can be allocated to the next DFD image.

It is also possible to specify a value of MaxAtoms either for individual frames or globally which is sometimes or always below the number of atoms that could be transmitted at the frame's bit budget. This is in effect a further rate control mechanism which would have the effect of reducing the bit rate of frames with a higher budget. For example this might occur if the motion vector cost is low.

5. EXPERIMENTS

The coding scheme is implemented within the H.264/JVC JM-92 Reference Code [5] to use its motion estimation and compensation module. The Intra frame coding method of H.264 is used, while all inter frames are coded with the proposed BUMP DFD coder. An IPPP frame structure is applied for coding the sequence, and 100 frames are coded each time, that is one Intra (I) frame and 99 Predicted (P) frames. All experiments are carried out on the luminance (Y) component of the Stefan and Hall sequences. Both are at CIF resolution (352 x 288 pixels), and the CABAC entropy coding mode is used.

The BUMP coder produces an embedded bit stream, so rate control can be easily implemented. A straightforward way to code a frame with a desired number of bits is to find an excessive number of atoms in the MP stage, called MaxAtoms, and code only as many as are required to meet the bit budget in the MERGE stage, unless all the available atoms have been coded. Our experiments show that MaxAtoms affects the average PSNR. This is because although MERGE is embedded, the MP atoms are not necessarily found in the same order that MERGE would code. In addition if MaxAtoms is small, all the Aatoms may be coded before the target bit rate is reached. The PSNR also depends on the quantization specified for the initial Intra frame whose coding is not embedded. With these two parameters are tuned for optimum performance, some comparisons are shown in Figures 3 and 4. Compared to H.263 as a baseline, along with the improvements introduced by the original Matching Pursuits codec of Neff and Zakhor [1], it is seen that both BUMP and H.264 give improved performance over a wide range of bit rates. For Stefan, BUMP is better than H.264 below 400K Bits/s and on Hall BUMP is always best.

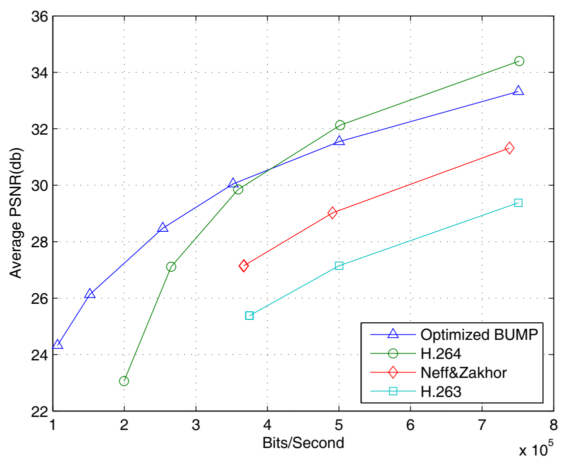

Figure 3. Rate/Distortion comparisons on Stefan.

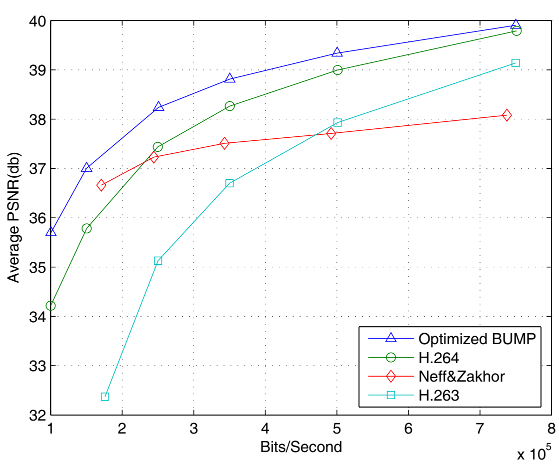

Figure 4. Rate/Distortion comparisons on Hall.

6. DISCUSSION AND CONCLUSIONS

A novel video coding scheme which incorporates a hybrid wavelet/matching pursuits coder for DFD frames into the H.264 reference software has been developed. Experiments show that this video coder can be very efficient, especially at low and medium low bit rates.

A strong feature of the method is that the DFD frames are coded in an embedded manner, unlike previous systems based on matching pursuits. This opens up a range of possibilities for rate control of the system from which further improvements may be expected. It is also likely that ongoing work on the run-length coder within MERGE could give bit rate reductions.

The experiments were limited to frame sequences consisting of an Intra frame coded by the non-embedded H.264 image coder, followed by a sequence of Predicted frames coded by our method. Future work will focus on the coding of Intra and Bidirectional frames, as well as more practical mixtures of frame types.

REFERENCES

[1] R. Neff and A. Zakhor, "Very low bit rate video coding based on matching pursuits", IEEE Trans. Circuits and Systems for Video Tech., vol. 7, pp. 158–171, 1997.

[2] Y. Yuan and D. M. Monro, "Improved Matching Pursuits Image Coding", IEEE Conf. Acoustics, Speech, and Signal Process. (ICASSP 2005), Philadelphia, March 2005.

[3] D. M. Monro, "Basis Picking for Matching Pursuits Image Coding", IEEE Int. Conf. Image Process. (ICIP 2004), Singapore, October 2004.

[4] D. M. Monro and Yuan Yuan, "Bases for low complexity matching pursuits image coding", IEEE Int.Conf. Image Process., Genoa, Sept. 2005.

[5] Karsten Sühring, H.264/AVC Reference Software, http://iphome.hhi.de/suehring/tml/

[6] D. M. Monro, J-L Aufranc, M. A.Bowers and W Poh, "Visual embedding of wavelet transform coefficients", IEEE Int. Conf. Image Process. (ICIP 2000), Vancouver, Sept. 2000.