MERGE CODING OF ATOMS FOR WAVELET/MATCHING PURSUITS IMAGE COMPRESSION

D. M. Monro, Xiaopeng Wang, Wei Huo and A. N. Evans
Department of Electronic and Electrical Engineering, University of Bath,Bath, BA2 7AY, United Kingdom
www.http://dmsun4.bath.ac.uk {d.m.monro, eepxw, eepwh, a.n.evans } @bath.ac.uk

ABSTRACT

Improved coding of atoms in image compression by Matching Pursuits (MP) after a wavelet decomposition is achieved. The positions of atoms in the wavelet sub-bands are communicated by run length coding. The MERGE coding scheme replaces individual coding of any alphabet of symbols by dividing the symbols into groups and run length coding their positions. It is shown theoretically that the use of MERGE with an efficient run length coder approaches the theoretical entropy of an alphabet of symbols. The embedded nature of MERGE makes it a near-ideal coding scheme for the amplitudes of atoms. An additional layer of MERGE coding for the basis indices of atoms approaches the theoretical optimum and is superior to either separable or combined Huffman coding.

Index Terms – Image coding, Wavelet transforms, Matching Pursuits, MERGE coding.

1. INTRODUCTION

This paper gives theoretical and practical details of the Multipass Embedded Residual Group Encoding (MERGE) coding scheme which is applicable to a wide range of lossless compression tasks. Applied to the coding of Matching Pursuits (MP) atoms, MERGE is shown here to be an effective and highly embedded coding scheme. MERGE groups symbols by value, and replaces the coding of a stream of symbols by a run length code of the positions of the symbols within the groups.

MP was introduced by Mallat and Zhang for digital audio [1], and applied by Neff and Zakhor to low bit rate video coding of motion compensated residual images [2]. In the authors' earlier work [3-5], gains in fidelity and reduced complexity were achieved by pretransformation by wavelets [3], embedded coding and improved dictionaries found by 'basis picking' [4, 5].

2. HYBRID WAVELET/MP CODEC

A hybrid wavelet/MP codec is used in this study, in which a multiscale wavelet decomposition is applied using a biorthogonal 9/7 filterbank before MP approximation. It has been shown that 5 scales for D1 (704 x 576) still images are a good choice [3]. Atoms for MP are then found on the 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 with the data set. In 2D it is usual to form the dictionary as a set of 1D bases applied separably. The code of an 'atom' consists of its position, sign, quantized amplitude and basis dictionary index. The atom is subtracted from the data and the process is repeated on the remaining residual. This is only a brief summary; MP is extensively documented in the literature [1-7].

Once a sufficient number of atoms has been found, for coding we raster scan the positions of the atoms out of the wavelet sub-bands from the highest to the lowest scale of decomposition, and in order LL, HL, LH, HH within each scale. For example, in Fig. 1 the order would be LL, HL2, LH2, HH2, HL1, LH1, HH1.

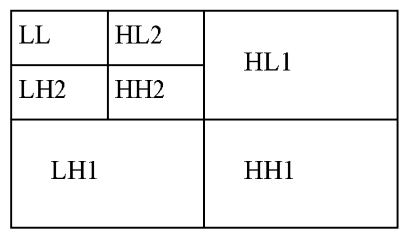

Fig. 1. Sub-band types after a 2 scale decomposition.

3. THEORY OF MERGE CODING

MERGE coding is a method we have used in our MP work for some time. This is the first detailed theoretical and practical justification of its effectiveness.

In many lossless compression tasks, a stream of symbols is to be communicated. Suppose a long stream of symbols is divided into groups A and B with independent stationary first order probability $p_B = p$ and $p_A = 1-p$. The expected cost of transmitting tokens to distinguish groups A and B is $-p\log_2 p-(1-p)\log_2(1-p)$ bits per token, so that the total cost of a transmission distinguishing the two groups by sending tokens for A and B by some perfect lossless code for a message of length N is $-N\{p\log_2 p+(1-p)\log_2(1-p)\}$ bits.

To MERGE code the grouping we send a Run Length Code (RLC) in which each token for group A is indicated by the number of tokens of B preceding it. For the first few possibilities for the RLC we get:

$p_A = 1-p$, $p_{BA} = p(1-p)$ and $p_{BBA} = p^2(1-p)$ and so on, giving the expected cost of a run length symbol in a message of infinite length as

$$C_A(\infty) = -(1-p)\log_2(1-p) - p(1-p)\log 2\{p(1-p)\}$$
$$-p^2(1-p)\log 2\{p^2(1-p)\} - \ldots \qquad (1)$$

which simplifies to

$$C_A(\infty) = -\frac{p}{1-p}\log_2 p - \log_2(1-p) \qquad (2)$$

For every token A, a run of tokens B is also defined. The rate of occurrence of A is $(1-p)$ so that the expected total cost of identifying both A and B is

$$E_{AB}(\infty) = (1-p)C_A(\infty) =$$
$$-(1-p)\left\{\frac{p}{1-p}\log_2 p + \log_2(1-p)\right\}$$
$$= -p\log_2 p - (1-p)\log_2(1-p) \qquad (3)$$

Therefore run length coding the occurrences of group A defines the positions of B also, and has exactly the same entropy as sending a stream of tokens for both A and B. This basic scheme extends to further division of groups A and B by making further RLC passes over the group to be divided. When coding multiple groups, the RLC should always skip positions in the message that are already known. If this is done it is easily shown that coding symbols α_i, $i = 1 \cdots N$ has entropy

$-\sum_{i=1}^{N} p(\alpha_i \log_2 p(\alpha_i))$, i.e. exactly the same as sending the symbols themselves by a perfect code.

In practice the length of the sequence of A and B tokens will be finite. Theoretical evaluation of equation (1) is more complicated for a finite sequence, but the entropy is reduced, so that Equation (3) is an upper bound. The gain in entropy of having a finite sequence is offset by having to signal the end of the groups. We find that using a special symbol for End of Group (EOG) is efficient except when many groups are empty.

4. MERGE CODING OF MP ATOMS

The coding of atoms could be done in a single pass, as was done by Neff and Zakhor [2], in which the space of the atoms is scanned, and the sign, amplitude and basis indices of the atoms are coded in some way, such as by a VLC. MERGE is an alternative that gives embedding of amplitudes and the possibility of accurate rate control. As seen in the previous section, in theory this is no less efficient a way of transmitting the amplitudes of atoms.

For our MP work we use Precision Limited Quantization (PLQ) of the amplitudes of atoms, which was originally motivated by psychovisual considerations [8]. In Fig. 2, if A is the amplitude of an atom before quantization, $S = sign(A)$, $FSB = \log_2|A|$, the First Significant Bit (FSB) of $|A|$, and R= the remaining bits in the range 0 to $2^{PL-1}-1$, for any precision limit $PL>1$, including the FSB.

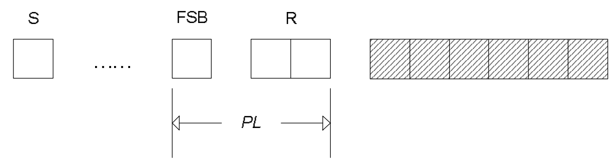

Fig. 2. Precision Limited Quantization of an atom of amplitude A by the triple $\langle S, FSB, R\rangle$.

Each atom chosen by MP occupies a position in the data space with the S, FSB and R attributes plus codebook index K. Lossless coding of all the attributes can be done by the MERGE algorithm, in which atoms are gathered into groups with all the attributes in common, and the positions are signalled by run length coding. Experimentally we have found this works well with PL=2, which helps to keep the number of groups small. To reduce the number of groups further, the sign S is coded as one bit of side information because positive and negative signs are equally probable. If the Precision Limit is PL, the full MERGE algorithm is:

For *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 (*FSB, R, K*).
    - Send the Sign S of the atom (1 bit)
  - End of Basis Function Group
- End of R (PLQ Residual) Group

End of *FSB* Group

5. CODING OF ATOM MAGNITUDES

In this section we investigate the effectiveness of MERGE for coding the amplitude residuals R of atoms. In MP it is found that the numbers of atoms of a particular quantized amplitude increases rapidly as the amplitude decreases. As the bit planes are scanned in decreasing order of *FSB,* the number of atoms increases rapidly. The amplitude residual R is one bit when PL=2, and there are always more atoms with R=0 than R=1 in a given bit plane, so that the entropy of R is less than 1 bit.

In Fig. 3 the amplitudes of atoms decrease from left to right as FSB planes are scanned. For each bit plane Fig. 3 shows the entropy of the R bits, which is always less than 1 because '0' is more common than '1'. The theoretical cost of MERGE taking into account the EOG symbols is also shown for each bit plane. Also shown is the actual cost of MERGE coding in a real codec, obtained by subtracting the RLC cost in bits of MERGE coding *FSB* and *R* from the cost of MERGE coding only *FSB* (without sending *K*). It is seen that the actual cost of MERGE approaches the theoretical cost closely as bit planes 14, 13 and 12 are reached. Clearly the VLC is not efficient because it cannot achieve a cost of less than 1 bit for *R*. In any practical application this means that MERGE is effective for coding *R* because the numbers of atoms required to code 'Gold Hill' between .1 bpp and .5 bpp are reached between FSB planes 14 and 12. A similar result is shown in Fig. 4 for the image 'Barbara 2.'

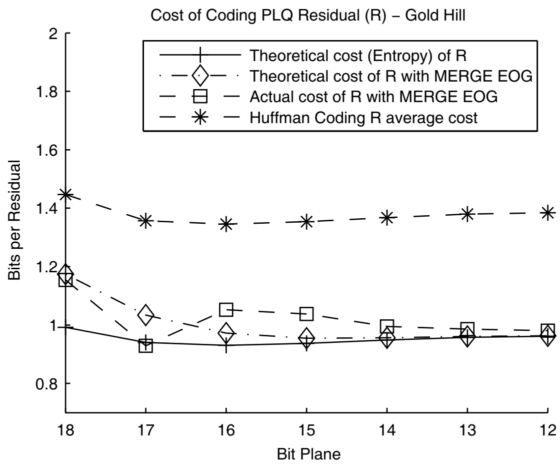

Fig. 3. Theoretical and actual costs of coding the residual bit *R* with *PL*=2, for the image 'Gold Hill'.

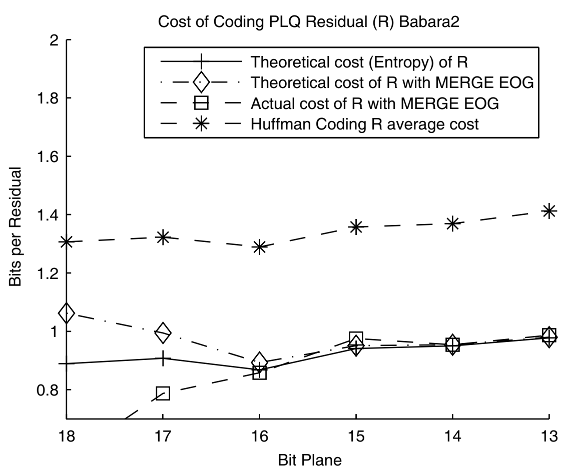

Fig. 4. Theoretical and actual costs of coding the residual bit *R* with *PL*=2, for the image 'Barbara 2'.

6. CODING OF BASIS INDICES

In this Section, MERGE coding of the basis indices *K* is compared with several VLC schemes. Neff and Zakhor [1] used an adaptive Huffman scheme in which the horizontal and vertical indices of the bases were coded separately. We compared this with the alternative of 'flattening' the 2D indices into a single 1D index – as is commonly done in computer programming when converting a 2D subscript to a 1D pointer. The result (not shown) is that the flattened approach works better.

In our most recent work we use separate codebooks for different types of sub-band [5], and only a sparse subset of the available 2D combinations is used. The flattening includes the removal of unused combinations. We investigated two versions of VLC for bases – one with sets four different adaptive Huffman tables for the four sub-band types, and another that uses only one adaptive Huffman table which refers to different indices in different sub-bands. One table works better over a wide range of bit rates, as illustrated by Fig. 5 for 'Gold Hill' and by Fig. 6 for the 'Barbara 2' image. In both cases it is seen that one adaptive Huffman table generally outperforms four, and over a wide range of bit rates MERGE coding the bases outperforms both VLC schemes.

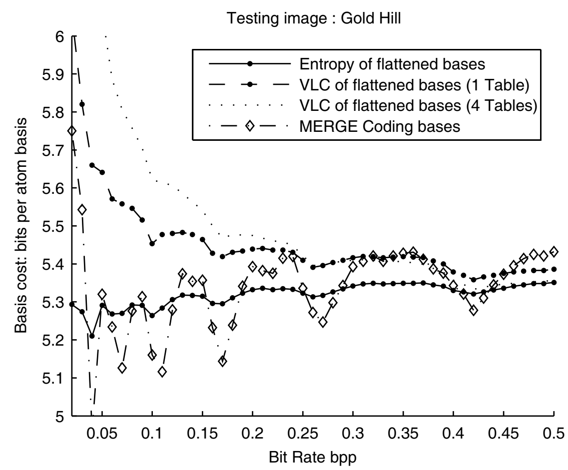

Fig. 5. Comparing Basis Coding schemes for 'Gold Hill'.

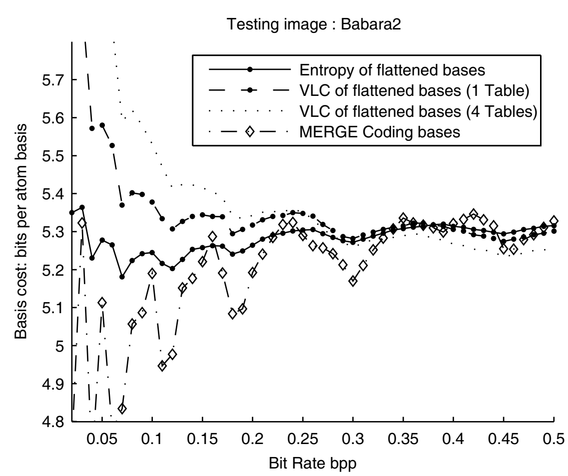

Fig. 6. Comparing Basis Coding schemes for 'Barbara 2'.

7. DISCUSSION AND CONCLUSIONS

It has been shown both theoretically and practically that MERGE coding is an effective, embedded method of coding atoms for MP. In Fig. 7 is shown a detailed Rate/Distortion graph for Gold Hill, showing the PSNR achieved relative to the JPEG 2000 Kakadu image coder Version 3.2. A full MERGE coding scheme is compared to two other MERGE schemes in which either the Basis indices are coded by a flattened 1D adaptive Huffman VLC scheme, or the Residual R bit of the atom code is sent as a single bit, as is done for the sign S. This has the advantage of reducing the number of EOG symbols sent in each FSB plane by a factor 2^{PL-1}. However the loss of embedding of the amplitudes is clearly a disadvantage, because mixing '0' bits with '1' bits for R generally underperforms MERGE. It is only at the ends of the *FSB* bit planes that sending raw bits for R gives comparable, and even occasionally better results. MERGE does not quite achieve 1 coded bit for R bit, but it is so close and the advantages of embedding are so great, that MERGE is the coder of choice for our hybrid wavelet/matching pursuits coding scheme. We have achieved an advantage over JPEG 2000 at low and medium bit rates, despite our MP scheme being of very low complexity, mainly as a result of our high performance basis dictionaries [4, 5]. Work will continue to improve high bit rate performance.

8. REFERENCES

[1] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries", IEEE Trans. Signal Processing, vol. 41, pp. 3397–3415, Dec. 1993.

[2] R. Neff and A. Zakhor "Very low bit rate video coding based on matching pursuits", IEEE Trans. Circuits Syst. Video Techol., vol. 7, pp. 158–171, Feb. 1997.

[3] Y. Yuan and D. M. Monro, "Improved Matching Pursuits Image Coding", Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., Philadelphia, March 2005.

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

[5] D. M. Monro, Wei Huo and Xiaopeng Wang, "Subband adaptive dictionaries for wavelet/matching pursuits image coding", IEEE Int. Conf. Image Process., Oct. 2006.

[6] P. Czerepinski, C. Davies, N. Canagarajah and D. Bull, "Matching pursuits video coding: dictionaries and fast implementation", IEEE Trans. Circuits Syst. Video Technol., vol. 10, No. 7, pp. 1103-1115, Oct. 2000.

[7] R. Neff and A. Zakhor "Matching Pursuit Video Coding - Part I: Dictionary Approximation", IEEE Trans. Circuits Syst. Video Technol., vol. 12, pp. 13-21, 2002.

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

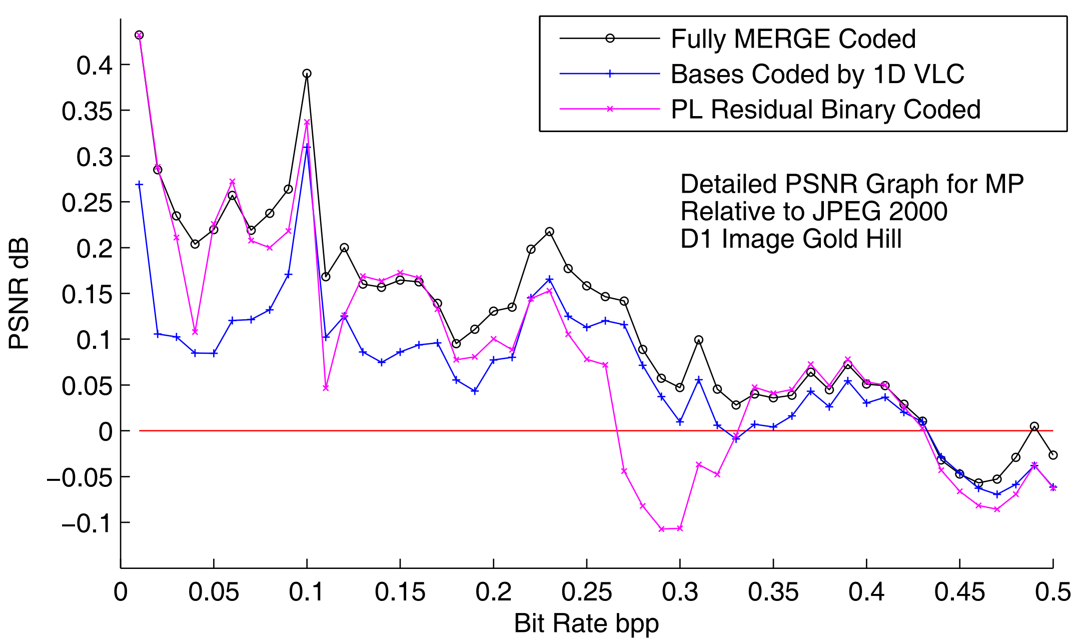

Fig. 7. Detailed R/D performance on Gold Hill for full and partial MERGE coding schemes applied to wavelet/matching pursuits still image coding with sub-band adaptive dictionaries.