A Separate Least Squares Algorithm for Efficient Arithmetic Coding in Lossless Image Compression

G. A. Triantafyllidis
and
M. G. Strintzis

Information Processing Laboratory
Aristotle University of Thessaloniki, Thessaloniki 540 06, Greece
Tel. : +3031.996351, Fax : +3031.996342
Email: strintzi@eng.auth.gr

ABSTRACT

The overall performance of discrete wavelet transforms for lossless image compression may be further improved by properly designing efficient entropy coders. In this paper a novel technique is proposed for the implementation of context-based adaptive arithmetic entropy coding. It is based on the prediction of the value of the current transform coefficient. The proposed algorithm employs a weighted least squares method applied separately for the HH, HL and LH bands of each level of the multiresolution structure, in order to achieve appropriate context selection for arithmetic coding. Experimental results illustrate and evaluate the performance of the proposed technique for lossless image compression.

Keywords: Least Squares, Arithmetic Coding, Context Selection, Lossless Image Processing

1 Introduction

Once an image has been decomposed, we need to entropy encode the transform coefficients. Two entropy coding methods are well-known and widely used: the Huffman and the arithmetic. The first method is preferable only when there is a lack of hardware resources and coding/decoding speed is a prime objective [Rab91]. Arithmetic is somewhat slower than Huffman, but it is much more versatile and effective. In most cases, the adaptive variant of arithmetic coding is used [Wil91],[Wit87], in order to take advantage from high order dependencies with the use of conditioning contexts.

The arithmetic data compression technique encodes data by creating code string which represents a fractional value on the number line between 0 and 1. On each recursion of the algorithm only one symbol is encoded. The algorithm successively partitions an interval of the number line between 0 and 1, and retains one of the partitions as a new interval. Thus, the algorithm successively deals with smaller intervals, and the code string lies in each of the nested intervals.

The performance of arithmetic coders depends mainly on the estimation of the probability model which the coder will use. The coder can achieve an average output code length very close to the entropy corresponding to the probability model it utilises. Therefore, if the probability model accurately reflects the statistical properties of the input, arithmetic coding will approach the entropy of the source. Different probability models will give different compression performance for the same data. Thus, a scheme with adaptive calculation of the probability of the data will be better than a non-adaptive scheme, as it will allow a better approximation to the "true" statistics of the data. The probabilities that an adaptive model assigns may change as each symbol is transmitted, based on the symbol frequencies seen so far in the message. A drawback of arithmetic coding of images using the above adaptive model is that it does not take into account the high amount of correlation between adjacent pixels. That is, each pixel is encoded using a probabilistic model adapted to all pixel values seen so far on the image. In this work, to alleviate this disadvantage a method similar to the one in [Wu98] is adopted, with which, for every new coefficient to be encoded, the model is updated more than once, making the probabilistic model more adaptive to recent pixels, and thus more effective.

Every transform coefficient is put into one of several classes (buckets) depending on the weighted values of a set of previously entropy coded coefficients. To each context type corresponds a different probability model and thus each subband coefficient is compressed with an entropy coder following the appropriate model. The key issue is then how to find an efficient context based classification.

In our work, the Magnitude-Set Variable-Length-Integer representation (proposed in [Sai96] and shown in Table 1) is employed to represent the transform coefficients. According to this, every coefficient is classified into one of a set of ranges called magnitude sets M, followed by the sign bit and the magnitude difference bits. For example, the numbers 15 and -16 are transmitted with the number triads $(7, +, 3)$ and $(8, -, 0)$ respectively.

This paper is organized as follows: Section 2 presents the manner for the tranform prediction and context selection. In Section 3 the weighted least squares error method is developed to determine the prediction of the current transform coefficient, in order to implement the context based adaptive arithmetic coding. Section 4 presents experimental results obtained when the proposed arithmetic entropy coder is applied, compared to S+P entropy coder. Finally, conclusions are drawn in Section 5.

Magnitude Set	Amplitude Intervals	Sign Bit	Magnitude Bits
0	[0]	no	0
1	[-1][1]	yes	0
2	[-2][2]	yes	0
3	[-3][3]	yes	0
4	[-5,-4],[4,5]	yes	1
5	[-7,-6],[6,7]	yes	1
6	[-11.-8][8,11]	yes	2
7	[-15,-12],[12,15]	yes	2
8	[-23,-16],[16,23]	yes	3
⋮	⋮	⋮	⋮

Table 1: Definition of Magnitude-Set Variable-Length-Integer $(MS - VLI)$ representation

2 Prediction and Context Selection

The magnitude set M of the current pixel is estimated using the weighted values of coefficients that have already been entropy encoded in the current band, in the sister band(s) and in the parent band, in the pyramid structure, i.e., the predictor has the form:

$$\hat{M} = \sum_{i=0}^{N} a_i M_i \qquad (1)$$

where the M_i indicates a previously encoded Magnitude Set, $\hat{M}$ is the prediction of the current Magnitude Set and the weights a_i, $1 \leq i \leq N$ are determined via linear regression so that $\hat{M}$ are least squares estimates of M.

Experimental results have proved that the magnitude sets of the coefficients shown in Fig. 1, which differ in shape for every subband LH, HL and HH, suffice for an accurate prediction of the magnitude set M of the current pixel. This scheme implies that the subbands will be coded in the following order: first the LH band, then the HL band and finally the HH band, so that to establish the necessary casual relationship. Therefore, Eq. (1) can be expressed for each of the subbands as follows [Tri99]:

$$
\begin{aligned}
LH\ band: \quad & \hat{M}_{LH}^k = a_1^k M_w^k + a_2^k M_{nw}^k \\
& + a_3^k M_n^k + a_4^k M_{ne}^k \\
& + a_5^k M_{p1}^k + a_6^k M_{p2}^k \\
HL\ band: \quad & \hat{M}_{HL}^k = b_1^k M_w^k + b_2^k M_{nw}^k \\
& + b_3^k M_n^k + b_4^k M_{ne}^k \\
& + b_5^k M_{p1}^k + b_6^k M_{p2}^k + b_7^k M_{sis}^k \\
HH\ band: \quad & \hat{M}_{HH}^k = c_1^k M_w^k + c_2^k M_{nw}^k \\
& + c_3^k M_n^k + c_4^k M_{ne}^k + c_5^k M_{p1}^k \\
& + c_6^k M_{p2}^k + c_7^k M_{sis1}^k + c_8^k M_{sis2}^k
\end{aligned} \tag{2}
$$

where k is the current level of decomposition, $a_i^k, i = 1, ..., 6,\ b_i^k, i = 1, ..., 7,\ c_i^k, i = 1, ..., 8$ are three groups of weights which are calculated separately via linear regression in order the $\hat{M}_x^k$ (where x =LH or HL or HH) are least squares estimate of M_x^k Subscripts w, nw, n, ne are directional short notations for west, north-west, north and north-east respectively, p_k $(k = 1, 2)$ indicates the k^{th} parent pixel and sis indicate the corresponding pixels or pixel in the sister bands.

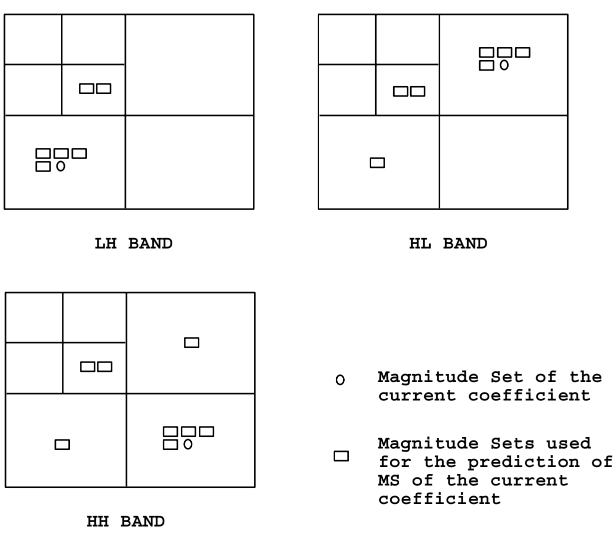

Figure 1: Pixels employed for the prediction of the magnitude set of current coefficient for each of the LH, HL and HH subbands

3 Optimization via appropriate Weighted Linear Regression

Let the matrices $\mathbf{S}_{LH}^k$, $\mathbf{S}_{HL}^k$ and $\mathbf{S}_{HH}^k$ have $L \times L$ rows, where L is the dimension of a band of the transformed image of the k level of decomposition and six or seven or eight columns respectively, depending on the specific band in which the coefficient belongs. Each row consists of all the previously encoded magnitude sets used for the estimation of the current magnitude set, i.e., the second subscript of each element of the matrices indicates the current coefficient while the second superscript indicates one of the six or seven or eight previously encoded magnitude sets used for the estimation of the current coefficient according to Eq. (2). Further, we form three vectors $\mathbf{y}_{LH}^k$, $\mathbf{y}_{HL}^k$ and $\mathbf{y}_{HH}^k$ composed of all magnitude sets of each band, i.e., the second subscript of each element of this vector indicates the current coefficient:

$$
\mathbf{S}_x^k = \begin{bmatrix}
M_{x,0}^{k,1} & M_{x,0}^{k,2} & \dots & M_{x,0}^{k,8} \\
M_{x,1}^{k,1} & M_{x,1}^{k,2} & \dots & M_{x,1}^{k,8} \\
M_{x,2}^{k,1} & M_{x,2}^{k,2} & \dots & M_{x,2}^{k,8} \\
\vdots & \vdots & \vdots & \vdots \\
M_{x,L\times L}^{k,1} & M_{x,L\times L}^{k,2} & \dots & M_{x,L\times L}^{k,8}
\end{bmatrix}
$$

$$
\mathbf{y}_x^k = \begin{bmatrix}
M_{x,0}^k \\
M_{x,1}^k \\
M_{x,2}^k \\
\vdots \\
M_{x,L\times L}^k
\end{bmatrix} \tag{3}
$$

where x =LH or HL or HH. Then, the vectors **a**, **b** and **c** of the optimal weights can be formed as [Cad90]:

$$
\begin{aligned}
\mathbf{a}^k &= (\mathbf{S}_{LH}^{k\ T} \mathbf{W}^k \mathbf{S}_{LH}^k)^{-1} \mathbf{S}_{LH}^{k\ T} \mathbf{W}^k \mathbf{y}_{LH}^k \\
\mathbf{b}^k &= (\mathbf{S}_{HL}^{k\ T} \mathbf{W}^k \mathbf{S}_{HL}^k)^{-1} \mathbf{S}_{HL}^{k\ T} \mathbf{W}^k \mathbf{y}_{HL}^k \\
\mathbf{c}^k &= (\mathbf{S}_{HH}^{k\ T} \mathbf{W}^k \mathbf{S}_{HH}^k)^{-1} \mathbf{S}_{HH}^{k\ T} \mathbf{W}^k \mathbf{y}_{HH}^k
\end{aligned} \tag{4}
$$

where $\mathbf{W}$ is the weighted linear regression matrix which may be chosen to be either the unity matrix ([Mem97], unweighted linear regression) or a user-defined appropriate weighted matrix ([Tri99], weighted linear regression).

Instead of using a simple unweighted linear regression algorithm, a more sophisticated method is implemented to find the best weights for the estimation of the current coefficient. More specifically, experiments have shown that the larger errors in estimating the transformed coefficients occur on the edges of the transformed image. As a result, the most appropriate matrix $\mathbf{W}^k$ of Eq. (4) must have higher weights in the positions which correspond to the edges detected on a band of the k level of decomposition of the transformed image. A Canny edge detector operator [Can86] is employed for this task.

Having calculated the weights, the norms of (2) can be used to classify the current transform coefficient to the proper bucket, that is to determine which probabilistic model to use during the adaptive arithmetic entropy coding.

In order to encode the LL band it can be decomposed again and similar techniques outlined above can be used to encode the high pass bands at this second level of decomposition. This procedure can be carried on until the low pass band is of very small size and can be transmitted in an uncoded manner, as is done in [Sai96].

The computation speed of the proposed method is somewhat slower than the simple arithmetic entropy coding since for each image that needs to be encoded, the calculation of the weights has to be performed in order to conclude to the fittest possible weights. However, if we want, we can go one step further: to use the linear regression algorithm for the weights calculation with a whole set of typical images. In that case, it is clear that compression performance is going to be decreased but the overall speed of the algorithm will be improved.

4 Experimental Results

The above context-based arithmetic entropy coding technique was compared to the method used in the widely regarded as state-of-the-art algorithm of S+P [Sai96]. Our experiments may be summarized as follows:

Step 1 Apply the S+P transform for the initial decorrelation of the selected image.

Step 2 Apply the algorithm of weighted least squares method to each band separately for the prediction of the magnitude sets.

Step 3 Classify the magnitude sets of each coefficient into one of several buckets depending on the weighted values of the selected set of previously entropy coded coefficients.

Step 4 Apply adaptive arithmetic entropy coding [Wil91] to each bucket. Aiming to better adaptivity, for every new coefficient to be encoded, the model is updated three times instead of once. The sign and magnitude difference are also arithmetically coded but using a fixed (instead of adaptive) uniform distribution model, in order to increase the computational efficiency.

The arithmetic entropy coder proposed, was applied to standard black and white, 8 bpp, images following an S+P transform [Sai96]. Table 2 presents the results of the unweighted or weighted least squares methods compared to that of the S+P entropy coder.

5 Conclusions

A method was presented for the implementation of an efficient context-based arithmetic entropy coding. The method employs weighted least squares techniques to determine the weights used so as to estimate the magnitude set of the current coefficient, based on a selected set of magnitude sets of pixels which have been previously coded. The set depends on the band in which the current transform coefficient belongs. Experiments show that the use of the weighted least squares algorithm to each band separately yields improved results, and consistently outperforms the entropy coder proposed by the state-of-the-art method S+P in [Sai96].

image	S+P		proposed method	
	bytes	bpp	bytes	bpp
lena	136702	4.17	135664	4.14
peppers	150182	4.58	149398	4.56
crowd	131010	4.00	130162	3.97
boat	141272	4.31	139909	4.27
airplane	128238	3.91	127020	3.88
bridges	182882	5.58	181027	5.52
harbour	154896	4.73	152902	4.67
barbara	149254	4.55	146931	4.48

Table 2: Number of bytes and bits per pixel needed for entropy coding with optimal weights calculated via linear regression with weighted matrix $\mathbf{W}^k$ (proposed method) compared to S+P entropy coding.

REFERENCES

[Rab91] M. Rabbani and P.W. Jones, "Digital Image Compression Techniques", Bellingham, WA: SPIE, 1991.

[Wil91] R. N. Williams, "Adaptive Data Compression", Norwell, MA: Kluwer, 1991.

[Wit87] R. M. Witten, I. H. Neal, and J. G. Cleary, "Arithmetic Coding for Data Compression", Communications of the ACM, vol. 30, pp 520-540, June 1987.

[Sai96] A. Said and W. A. Pearlman, "An image multiresolution representation for lossless and lossy compression", IEEE Transactions on Image Processing, vol. 5, pp 1303-1310, Sept. 1996.

[Yoo96] Y. Yoo, A. Ortega, and B. Yu, "Adaptive Quantization of Image Subbands with Efficient Overhead Rate Selection", Proc. of the Intl. Conf. On Image Proc., ICIP96, vol. 2, (Lausanne, Switzerland), pp 361-364, Sept. 1996.

[Mem97] N. Memon, X. Wu, and B.-L. Yeo, "Entropy Coding Techniques for Lossless Image Compression with Reversible Integer Wavelet Transforms", IBM Research Report, Computer Science/Mathematics, RC 21010, Oct. 22, 1997.

[Wu97] X. Wu and N. D. Memon, "Context-Based Adaptive Lossless Image Coding", IEEE Transactions on Communications, vol. 45, pp 437-444, Apr. 1997.

[Wu98] X. Wu, K. U. Barthel and G. Ruhl, "Adaptation to Nonstationarity of Embedded Wavelet Code Stream", Proc. of the Intl. Conf On Image Proc., ICIP98, (Chicago, IL), Oct. 1998.

[Can86] J. Canny, "A computational approach to edge detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 8, No. 6, pp 679-698, Nov. 1986.

[Cad90] J. A. Cadzow, "Signal Processing via Least Squares Error Modeling", IEEE Magazine of ASSP, pp 12-31, Oct. 1990.

[Efs96] S. Efstratiadis, D. Tzovaras, and M. G. Strintzis, "Hierarchical Partition Priority Wavelet Image Compression", IEEE Transactions on Image Processing, Vol. 5, No. 7, pp 1111-1123, July 1996.

[Tri99] G. A. Triantafyllidis, M. G. Strintzis, "A Context Based Adaptive Arithmetic Coding Technique for Lossless Image Compression", IEEE Signal Processing Letters, Vol. 6, No. 7, pp 168-170, July 1999.

[Tri00] G. A. Triantafyllidis, M. G. Strintzis, "A Least Squares Algorithm for Efficient Context-Based Adaptive Arithmetic Coding", submitted to ISCAS 01, (Sydney Australia), May 2001.