A LOSSLESS COMPRESSION ALGORITHM BASED ON PREDICTIVE CODING FOR VOLUMETRIC MEDICAL DATASETS

M. Díez-García, F. Simmross-Wattenberg, C. Alberola-López

Laboratorio de Procesado de Imagen, E.T.S.I. Telecomunicación,
Univ. Valladolid, 47011, Valladolid, Spain
email: {mdiegar}{fedesim}{caralb}@lpi.tel.uva.es
web: www.lpi.tel.uva.es

ABSTRACT

In this paper, a novel lossless compression method for volumetric medical datasets based on 3D adaptive prediction is presented. While almost all existing methods are based on a three-dimensional wavelet transform, this method, referred to as LSE-3D, evaluates the performance of an algorithm that operates on a least-square prediction basis. This leads to a different method with approximately similar results on compression performance, although with some advantages on computational cost.

1. INTRODUCTION

An important problem in medical imaging is that of efficient volumetric image compression, since the amount of medical data would easily overwhelm the storage and transmission systems. Lossless compression is preferred by physicians in order to avoid many legal and regulatory issues [1]. There is a broad range of medical image sources, and for most of them discarding small image details that might be an indication of pathology could alter a diagnosis, causing severe human and legal consequences.

General lossless compression engines are considered to be composed of two main operational blocks: a data decorrelation block and a second stage aimed at an entropy codification of decorrelated data. Usually, two major tendencies have been distinguished on the decorrelation module: former employs wavelet transforms, whereas the latter supports methods based on predictive coding. This fact is clearly reflected in principal ITU-T compression standards: JPEG2000 [2] is the main representative of the first group, while JPEG-LS [3] is the best paradigm of the latter.

The increasing use of three-dimensional imaging modalities, like Magnetic Resonance Imaging (MRI), Computerised Tomography (CT), Ultrasound (US), Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET), triggers the need for efficient techniques to transport and store the related volumetric data. In this context, it is meaningful that the ISO/IEC JTC1/SC29/WG1 committee decided to develop an extension (i.e. Part 10) of JPEG2000 that will give support to three-dimensional encoding mechanisms [4]. Recently several examples of 3-D wavelet-based coding engines have arisen, either from direct extensions to a third space of already defined still image compression methods, like 3D SPIHT (Set Partitioning in Hierarchical Trees) [5], or just defined as new algorithms, like CS-EBCOT (Cube Splitting) [6] or 3D QT-L (QuadTree-Limited) [7]. Usually these techniques are rich featured but computationally intensive, with typical bottlenecks in memory access [6, 8]. While intensive attention is paid to transform-based compression methods, the problem of adaptive prediction coding applied to volumetric image compression is relatively under-investigated.

It has been proved for still continuous-tone images [9] that JPEG-LS offers quite an efficient performance, since compression rates obtained have reached very similar results to those of JPEG2000 in spite of its greater simplicity, both computational and conceptual. Therefore, it is reasonable to think that an extension of basic JPEG-LS premises to 3D images shall have similar satisfactory results. This paper focuses on this aspect, trying to shed light on a compression algorithm computationally efficient.

Predictive coding methods are founded on a coding process of residuals together with the prediction model employed, instead of directly perform image data coding. Prediction errors, which hopefully shall have small amplitudes with fairly high probability, are coded using entropy coding techniques that tend to associate shorter binary codes to most probable symbols. Compression is achieved this way, being lossless whenever prediction model is spatially causal. In this context, the most efficient schemes employ a context-based entropy coding, i.e., a different entropy coder is used depending on the values of adjacent pixels. Since a context value describes the local properties of the volume region where a prediction error occurs, a more accurate description of the prediction error distribution is obtained within the context, and therefore higher compression rates are achieved.

Linear prediction is an efficient decorrelation tool for stationary sequences. Although it is commonly assumed that natural images are characterized by abrupt changes in local statistics, the motivation behind linear adaptive prediction is based on fact that image values can be considered as locally stationary sequences, so it would be possible to optimally exploit dependencies within causal context of the pixel by an adaptive model. Context-based adaptive prediction schemes [10], like those implemented in JPEG-LS and CALIC, have proved to achieve significant improvements over fixed predictors such as lossless mode of JPEG standard. They can be seen as variable prediction models that change according to a experimentally tuned switching function, to adapt to local statistics. In this sense, least-square(LS) adaptive linear prediction schemes [12] have demonstrated important improvements over former models due to their natural adaptation to actual data.

The rest of the paper is organized as follows. Section 2 presents the proposed prediction-based compression scheme for three-dimensional images. The lossless coding results obtained for five volumetric data sets recorded with different imaging modalities are shown in section 3, together with a comparative analysis between our coder and some wavelet-transform based methods. Finally, section 4 summarizes the conclusions.

2. LSE-3D CODER DESCRIPTION

In this section, we propose a lossless compression scheme for volumetric images, referred to as LSE-3D (Least Square Estimation in 3-D), which is characterized with improved efficiency compared with state-of-the-art adaptive prediction-based algorithms. It is based on a scheme for still image coding presented in [11], extending it in order to consider information from third spatial dimension.

The authors acknowledge the Comisión Interministerial de Ciencia y Tecnología, Spain, for research grants TIC2001-3808-C02-02 and TEC2004-06647-C03-01. Acknowledgments are extended to the European Commission for the funding associated to the Network of Excellence SIMILAR (FP6-507609). Special thanks are also given to Dr. Schelkens for providing the experimental data and to Mr. Rafael Redondo for his support on wavelet coders evaluation .

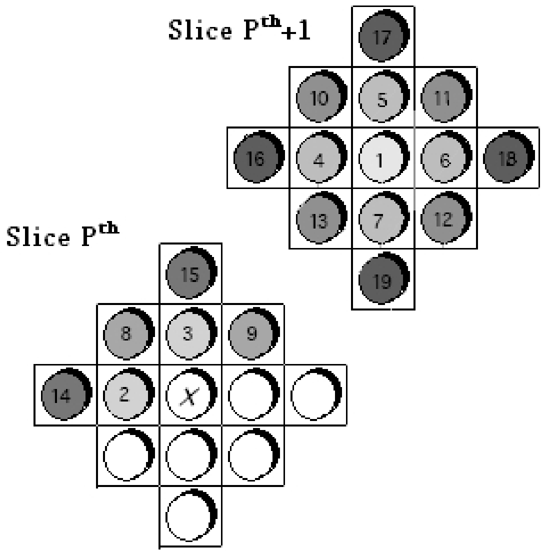

Figure 1: Ordering of 3-D causal context, based on euclidean distance to current voxel.

In other words, we will try to remove spatial redundancy present in all three dimensions of volume, taking into consideration the values of adjacent voxels, with low delay and limited complexity.

2.1 Least-square linear prediction

A reasonable assumption made with natural images is the Nth order Markovian property. That is, we only need to consider the N nearest causal neighbors in the prediction, that will shape the ”‘context”’ of the prediction.

$$\hat{S}(i) = \sum_{k=1}^{N} \alpha_k S(i-k)$$

In order to explain proposed LS-based prediction method in a easy and intuitive manner, an order relationship established among context voxels will be defined. A 3-D euclidean distance is defined as

$$d_{3-D} = \sqrt{(m-m_o)^2 + (n-n_o)^2 + (p-p_o)^2}$$

where $S(m,n,p)$ is the voxel in causal context of $S(m_o,n_o,p_o)$, that represents processed voxel. Usually, volumetric images will be scanned in a raster-scanning order along each slice. This way, it will be possible to adopt i to denote the spatial coordinate, being $S(i)$ the value of the voxel in coordinates (m,n,p). A three-dimensional causal neighborhood could be defined over current and previous slice, according to an ordering such as the one shown in Figure 1, allowing us to refer to the kth nearest neighbor in the causal context as $S(i-k)$.

The LS approach enables us to find the optimal coefficients a_k such that the sum of squares of differences between the prediction $\hat{S}(i)$ and the actual value $S(i)$ is minimized. This will allow to achieve local statistics adaptation and obtain the optimum prediction within a causal region (called ”‘training window”’). Denoted by $\vec{x} = [S(i-1)\cdots S(i-M)]^T$, this window will describe the local region where the LS-based estimator is being assessed. Translating this training window to each component of the linear prediction we would obtain an $M \times N$ matrix

$$C = \begin{bmatrix} S(i-1-1) & \cdots & S(i-1-N) \\ \vdots & & \vdots \\ S(i-M-1) & \cdots & S(i-M-N) \end{bmatrix}$$

where $S(i-m-n)$ is the mth prediction neighbor of $S(i-n)$. Using standard theory on optimal linear prediction, results from the minimization of the prediction error power are derived from

$$\vec{\alpha} = (C^T C)^{-1} C^T \vec{x}.$$

It must be noted that $\vec{\alpha}$ and $\hat{S}(i)$ are computed on the basis of past information, so the decoder would be able to replicate all operations without requiring any side-information. Coefficient assessment therefore implies the calculation of the inverse of a square matrix of size N. This can be done using the Gauss-Jordan elimination process, but considering that $C^T \cdot C$ represents the covariance matrix of the considered voxel values, and that it is consequently definite positive, the inverse could be effectively assessed as well by Cholesky decomposition, with half computational cost [13].

Initially, coefficients assessment should be done for each voxel in volume, in order to achieve maximum local statistics adaptation. But in practice, this would derive in a great and often inefficient computational load. Thus, in same manner as in [11], we will consider that while the prediction error does not exceed a predefined threshold T linear coefficients will not be recalculated, allowing to perform LS optimization only when necessary. Furthermore, whenever covariance matrix is singular, coefficients will be initialized to $1/N$, thus assuming that the prediction is the mean value of the considered context.

2.2 Error prediction coding

Once the prediction error has been obtained (as the floor value of the difference between the present voxel value and the assessed prediction), this error is entropy coded. Continuing with our philosophy of local statistics adaptation, a context-based code will be used; actually a Golomb-Rice code, similar to that defined in JPEG-LS standard [3], will be proposed due to its simplicity and relatively low computational cost.

The prediction error, e, is an integer random variable which is entropy coded using a context based probabilistic model. It can be represented with no loss of information using a modulo-2^b representation in the range $[-2^{b-1}, 2^{b-1}-1]$, where b is the bit precision. This is because any representation of the form $\hat{e} = e + k \cdot 2^b$ permits to recover $S = e + k \cdot 2^b + \hat{S}$ with no ambiguity, since the decoder can compute $\hat{S}$ and it is known in advance that S is in the range $[0, 2^b - 1]$. Thus, there is no need for an extra bit in the representation of the difference $e = S - \hat{S}$. The modulo-2^b representation of e typically has a two-sided geometric distribution, possibly with a nonzero average value which is actually estimated and removed before coding, via a bias correction procedure. Furthermore, to map the two-sided geometric distribution into a one-sided one, for which computationally efficient coding schemes exist, positive and negative values of the prediction error are interleaved and remapped into the range of non-negative integers. Finally, a Golomb-Rice code is used to represent these prediction residuals.

Golomb-Rice codes are a family of codes indexed by a single parameter, which clearly depends on the input distribution characteristics. Like JPEG-LS, our proposal does not use a single probabilistic model for the prediction error distribution, but a set of 365 models and, consequently, of Golomb-Rice codes. Each model is

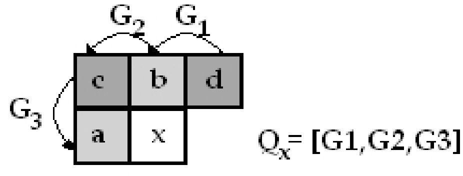

Figure 2: Context considered in prediction residual coding.

Name	Resolution	Bit Depth
ECHO	256x256x256	8 bpp
MR	256x256x200	12 bpp
CT	512x512x44	12 bpp
PET	128x128x39	15 bpp

Table 1: Test images used in performance analysis.

chosen within the particular context in which the prediction error has occurred. The context value is calculated by quantizing a vector of pixel differences, namely $[d-b, b-c, c-a]$(see Figure 2). For each context, the prediction error distribution is estimated using errors in the past, and then the appropriate Golomb-Rice code is selected. JPEG-LS also includes a run-mode procedure to code image regions with constant pixel values, which is maintained in our proposal. Thus, when processing a low entropy region, with null context value, the algorithm will be further simplified as only the run-length value will be coded.

3. EXPERIMENTAL RESULTS

In this section, some experimental results are reported, evaluating the performance corresponding to different options for the parameters of the proposed lossless compression algorithm. Several test were done on four volumes, representative of most-used modalities of medical imaging: a magnetic resonance (MR), an echography (ECHO), a positron-emission tomography (PET), and an X-ray computerized tomography (CT). These volumes were also used in [6] for evaluation and will allow us both to evaluate our algorithm with different image sizes and bit depths and to compare it with other state-of-the-art techniques, in order to achieve a greater validity for our compression algorithm.

3.1 Performance analysis

Lossless coding results are reported for several implementations of our LSE-3D coding method, using a set of values for M, N and T parameters on each of selected test volumes. For completeness in the performance study, these results need to be compared against those achieved by state-of-the-art techniques, like 3D-SPIHT [5] or 3D QT-L [6]. Obviously, just providing an objective measure for compression performance will not be enough for a fair comparison. Hence, we also analyze the computational cost associated to each implementation of the considered techniques.

Since measure dependency on image properties was not desired, compression efficiency was chosen for doing the performance comparison, computed as

$$CE = \frac{Orig.\ Image\ Size - Compressed\ Image\ Size}{Orig.\ Image\ Size} \times 100$$

.

Figure 3 shows the lossless coding efficiency results achieved by several implementations of the LSE-3D coder, together with the results (reported in [6]) achieved by two wavelet-based techniques, taking the latter as reference values. Denoted by LSE-3D(M,N), reported implementations are provided with N-th order linear predictors, where linear coefficients assessment is done over M-size contexts. Normally, values of parameter M are chosen around N^2 for better performance [11], while parameter T, that represents the error threshold that triggers the assessment of linear coefficients, is bit-depth dependent.

Figure 4 reports computational costs in terms of floating-point operations (sums and multiplications) per voxel derived from execution of LSE-3D prediction algorithm on each test volume. We can infer a clear compromise between this magnitude and compression performance that must be solved in each case.

We notice that for all volumes, the LSE-3D coder achieves an similar coding performance than wavelet-based schemes for low values of T, providing even better results in the ECHO and PET

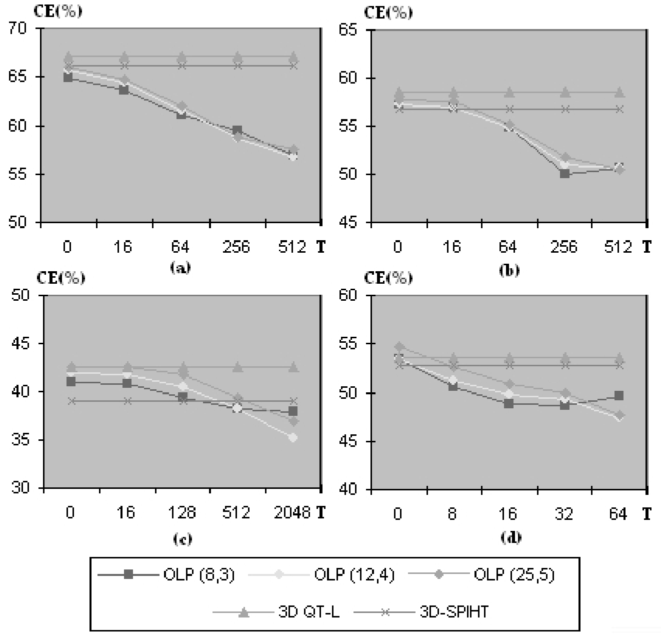

Figure 3: Compression efficiency (CE) derived from evaluated lossless coders implementations over test volumes, (a) MR (b) CT (c) PET (d) ECHO. LSE-3D coder results are reported for several values of T parameter.

volumes. The compression efficiency clearly gets diminished with greater values of threshold T, although at the same time the number of operations per voxel noticeably drops to very low levels. Since LSE-3D is designed for greater performance in positions with limited computational resources, gain in computational cost will prevail over compression efficiency. Therefore the selection of the values for the coder parameters in each volume will be submitted to the achievement of a good compromise between these performance measures.

3.2 Comparative analysis

Due to the different nature of our proposed method and the wavelet-based coders considered, any fair comparison should regard not only measurable performance but also any additional features supported. For example, both 3D QT-L and 3D-SPIHT have resolution and quality scalability because of their use of wavelet transforms and progressive coding algorithms. In this sense, we will compare the complexity of the decorrelation method separately from the one derived from the codification module in both compression strategies, i.e., the residual prediction coding or the wavelet-based techniques.

Regarding codification, our method based on Golomb-Rice codes was aimed to be simple yet efficient, with little computational resources. From this point of view, it clearly outperforms other coding methods employed in wavelet-based techniques, where usually complex optimization algorithms and arithmetic codes are used. However, due to the fact that these coding processes carry out a set of complementary features that our codec does not support, i.e., progressive coding, we will take the complexity of the coding algorithm as a consequence of the further desirable characteristics of the compression method.

Thus far, decorrelative processes stand out as decisive in comparative analysis. Figure 4 illustrates the number of floating-point operations (sums and multiplications) per voxel associated to a 5-level wavelet transform that employs a lossless $5x3$-lifting kernel in all 3 dimensions, together with those carried out during prediction algorithm in each LSE-3D implementation. The former slightly

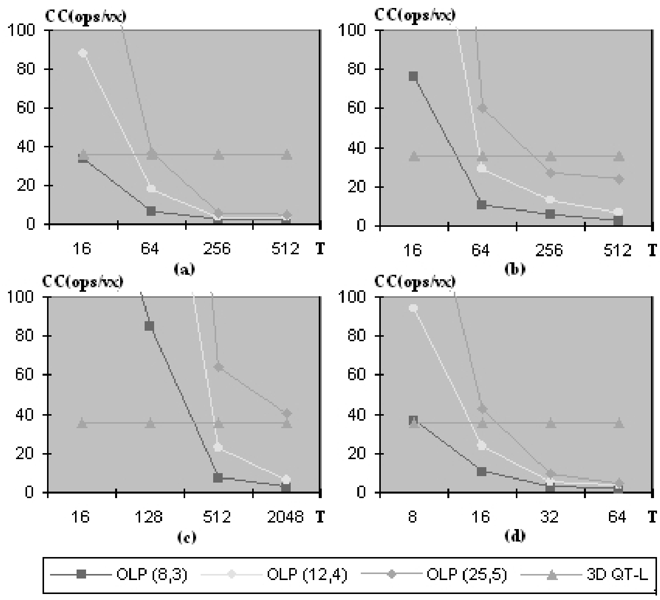

Figure 4: Computational cost (CC) derived from evaluated lossless coders implementations over test volumes, (a) MR (b) CT (c) PET (d) ECHO. LSE-3D coder results are reported for several values of T parameter.

vary around 36 operations per voxel, whereas the latters are strongly dependent of T parameter. It can be noticed that, for all volumes, an LSE-3D coder implementation delivers up to 50% less computational load than wavelet-based algorithms, while provided compression efficiency is maintained within a 5% interval of the best performance.

4. CONCLUSIONS

This paper proposes a new lossless coding algorithm based on a predictive scheme aimed to be simple but efficient, referred to as LSE-3D. Its performance was compared with state-of-the-art techniques, including 3D SPIHT and 3D QT-L, analyzing both compression efficiency and computational load. Based on a test bed of 4 medical volumes it was shown that our LSE-3D coder performs a bit below wavelet-based techniques, although its computational requirements are much less restrictive. This work provide competitive lossless coding results on positions with critical restrictions imposed on available computational resources.

REFERENCES

[1] S. Wong, L. Zaremba, *et al.*, "Radiologic Image Compression - A Review", *Proceedings of the IEEE*, vol. 83, no. 2, pp. 194-219, Feb 1995.

[2] ISO/IEC 15444-1, *Information Technology - JPEG 2000 image coding system*, JPEG 2000 Standard, Part 1- Core coding system, 2000.

[3] ISO/IEC 14495-1, *Information Technology -Lossless and near-lossless compression of continuous-tone still images*, JPEG-LS Standard, Baseline, 2000.

[4] C. Brislawn and P. Schelkens, "JPEG 2000 Part 10: Extensions for Three-Dimensional and Floating Point Data - Scope and Requirements", ISO/IEC JTC1/SC29/WG1, Report WG1N2378, 2001.

[5] Y. Kim and W. A. Pearlman, "Stripe-based SPIHT lossy compression of volumetrical medical images for low memory usage and uniform reconstruction quality," in *Proc. ICASSP*, Istanbul, Turkey, June 5-9. 2000, vol. 4, pp. 2031-2034.

[6] P. Schelkens, A. Munteanu, *et al.*, "Wavelet Coding of Volumetric Medical Datasets", *IEEE Transactions on Medical Imaging*, vol.22, no.3, March 2004.

[7] A. Munteanu, J. Cornelis, *et al.*, "Wavelet-based lossless compression scheme with progressive transmission capability", *International Journal of Imaging Systems and Technology*, vol. 10, no. 1, pp. 76-85, Jan 1999.

[8] K. Varma and A. Bell, "JPEG2000-Choices and Tradeoffs for Encoders", *IEEE Signal Processing Magazine*, pp.70-74, Nov 2004.

[9] D. Santa-Cruz, T. Ebrahimi, *et al.*, "JPEG 2000 still image coding versus other standards", *Proc. SPIE's 45th annual meeting, Applications of Digital Image Processing XXIII*, San Diego, California, Jul 30-Aug 4. 2000, vol. 4115, pp. 446-454.

[10] N. Memon and X. Wu, "Recent developments in context-based predictive techniques for lossless image compression",*Computer Journal*, vol. 40, no.2/3, pp. 127-136, 1997.

[11] X. Li and M. Orchard, "Edge-directed prediction for lossless compression of natural images",*IEEE Trans. Image Processing*, vol. 10, no.6, pp. 813-817, 2001.

[12] H. Ye, G. Deng and J. Devlin, "Least squares approach for lossless image coding",*Proc. ISSPA*, Brisbane, Qld., Australia, Aug 22-25. 1999, pp. 58-62.

[13] J. Makhoul, "Linear prediction: a review", *Proc. IEEE*, vol. 63, pp. 561-580, Feb 1975.