SERBIAN JOURNAL OF ELECTRICAL ENGINEERING
Vol. 8, No. 1, February 2011, 27-36

UDK: 004.92.032.2:616-7

Lossless Predictive Compression of Medical Images*

Aleksej Avramović[1], Slavica Savić[1]

Abstract: Among the many categories of images that require lossless compression, medical images can be indicated as one of the most important category. Medical image compression with loss impairs of diagnostic value, therefore, there are often legal restrictions on the image compression with losses. Among the common approaches to medical image compression we can distinguish the transformation-based and prediction-based approaches. This paper presents algorithms for the prediction based on the edge detection and estimation of local gradient. Also, a novel prediction algorithm based on advantages of standardized median predictor and gradient predictor is presented and analyzed. Removed redundancy estimation was done by comparing entropies of the medical image after prediction.

Keywords: Medical Imaging, Digital image processing, Lossless image compression, Prediction.

1 Introduction

Digital radiology has resulted in significant increase of use of digital medical images in the process of diagnosis [1]. In order to preserve the value of diagnostic medical images, it is necessary to provide lossless image compression. Apart from practical reasons, there are often legal restrictions on the lossless medical image compression. As for the method of compression, predictive compression is much simpler then transformation-based compression, and in addition usually results in lower bit rate. During recent years, several algorithms for predictive lossless image compression have been presented.

Predictive algorithms for image compression can be classified in two groups:

- the algorithms with a single pass, and
- the algorithms with two passes.

Algorithms with one pass count all required parameters for compression during one scanning of image, such as the optimal predictor parameters. Coding

[1]Faculty of Electrical Engineering, University of Banja Luka, Patre 5, 78000 Banja Luka, Bosnia and Herzegovina, E-mails: aleksej@etfbl.net; slavica.savic@etfbl.net

*Award for the best paper presented in Section *Electric Circuits and Systems and Signal Processing*, at Conference ETRAN 2010, June 7-11, Donji Milanovac, Serbia.

of prediction error can be done in combination with prediction or in a separate phase after the prediction. Algorithms with two passes have the ability to analyze image and efficiently determine the optimal parameters for the prediction and coding in the first pass, while in the second pass consist of prediction and coding. Optimization of parameters is usually based on the principle of least square error, which generally gives better results than switching predictors. Switching predictors choose sub predictor from the set of predictors based on the context in which the pixel is located.

This paper is organized as follows. The second section gives a brief description of the predictive methods of compression, while the third section describes the most important predictors used in the proposed algorithms. The proposed predictor and its characteristics is also described. The fourth section provides a comparative analysis of the proposed predictor with other predictors, while the fifth section is conclusion.

2 Predictive Image Compression

Predictive compression is based on several stages, as shown in Fig. 1, that are: prediction, contextual modeling, error modeling and entropy coding. This type of lossless method of compression became a standard during the last decade of the last century. At the time, several algorithms were proposed and tested for the purpose of adoption of a standard for lossless image compression [2]. Prediction is the crucial part of the compression, because it removes most of the spatial redundancy, and the choice of the optimal predictor is essential for the efficiency of compression methods. The prediction may be linear or nonlinear. Linear prediction is based on one or more fixed predictors, which can be combined with appropriate weights or it is possible to choose the optimal sub predictor on the basis of the properties of a certain part of the picture, i.e. on the basis of context of the region. The context can be determined for each pixel separately, for example, during the raster scanning, pixel can be found in the region of the horizontal edge, vertical edge, sloping edges, certain texture, smooth region, etc. Linear predictors based on the finite group of sub predictors, are simple and fast, a significant advantage is possibility of realization of the integer system. Several predictors, based on the least squares optimization, have been proposed, which in comparison with fixed predictors can be more efficient, but they are much more complicated and computationally demanding. Optimization is based on finite group of casual pixels, so called, context. Since the optimization for each pixel can be computationally demanding, adaptation of coefficients is usually done when another type of region occurs, for example an edge. For effective adaptation a larger context, comparing with fixed predictors, is required. Nonlinear prediction is based on neural networks, vector quantization, etc.

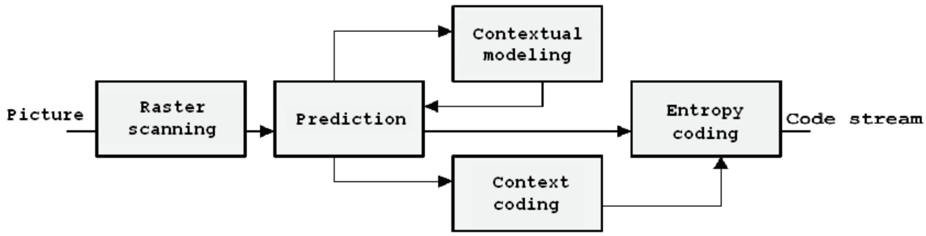


Fig. 1 – *General scheme of lossless predictive image compression.*

Contextual modeling means adaptive correction of prediction of pixels in order to exploit repeated schemes in a picture. For example, for certain layout of values of contextual pixels is possible to use information about prediction for the next occurrence of such layout and correct prediction for each time these occur. Even for the 8-bit images, the number of contexts, defined according to the layout of pixel values, significantly increases with increased the number of contextual pixels. Therefore, a much smaller number of contexts are defined. Particular context may be selected using quantization of contextual pixel values or using other transform operation.

Modeling error of prediction enables more efficient coding using entropy coder. Error modeling can further reduce the entropy of prediction error image. Mathematically, the conditional probability of the error prediction in relation to the particular context of the error is smaller compared to the probability of errors without contexts. In addition, it is possible to transform the alphabet properly adapted to a particular type of entropy coder. For example, there are Golomb-Rice versions of codes that require non-negative input data.

Entropy coding removes statistical redundancy of the prediction error images. The most commonly used are Huffman code, Golomb-Rice code, or arithmetic coding. Entropy coding actually performs the compression and gives the final output sequence of compressed data. Coding can be done after the prediction as an independent phase or progressively during the prediction.

3 Linear Predictors

As the most important predictors for lossless image compression the median predictor used in standard JPEG-LS and the gradient predictor, used in CALIC algorithm, are emphasized.

Median Edge Detection (MED) belongs to the group of switching predictors, that select one of the three optimal sub predictors depending on whether it found the vertical edge, horizontal edge, or smooth region [3]. In fact, MED predictor selects the median value between three possibilities *W*, *N* and $W + N - NW$ (common labels are chosen after sides of the world, Fig. 2), and the optimal combination of simplicity and efficiency. The prediction is based on:

$$P = \begin{cases} \min(W,N), & if\ NW \geq \max(W,N), \\ \max(W,N), & if\ NW \leq \min(W,N), \\ N + W - NW, & else. \end{cases} \quad (1)$$

Gradient Adjusted Predictor (GAP) is based on gradient estimation around the current pixel [4]. Gradient estimation is estimated by the context of current pixel, which in combination with predefined thresholds gives final prediction. GAP distinguishes three types of edges, strong, simple and a soft edge, and is characterized by high flexibility to different regions. Gradient estimation is done using:

$$\begin{aligned} g_v &= |W - WW| + |N - NW| + |N - NE|, \\ g_h &= |W - NW| + |N - NN| + |NE - NNE|, \end{aligned} \quad (2)$$

and the prediction is made by the algorithm:

$$\begin{aligned} & if\ g_v - g_h > 80, P = W \\ & elseif\ g_v - g_h < -80, P = N \\ & else\ \{P = (W + N)/2 + (NE - NW)/4 \\ & if\ g_v - g_h > 32, P = (P + W)/2 \\ & elseif\ g_v - g_h > 8, P = (3P + W)/4 \\ & elseif\ g_v - g_h < -32, P = (P + N)/2 \\ & elseif\ g_v - g_h < -8, P = (3P + N)/4\}, \end{aligned} \quad (3)$$

where the label of causal pixels in (2) and (3), also harmonized according to Fig. 2.

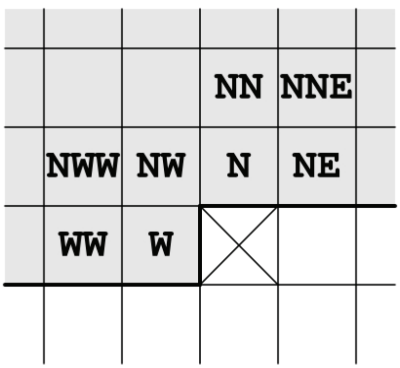


Fig. 2 – *Causal and contextual neighbors.*

The authors in [5] have proposed adaptive prediction based on a combination of thirteen simple predictors and the appropriate penalty of predictors which result in large prediction errors. In the same work, similar results can be obtained using only six simple predictors. The six predictors defined in a P6 set, are:

$$\begin{gathered} W, \quad W+N-NW, \quad N, \\ NE, \quad (NW+W)/2, \quad NW, \end{gathered} \tag{4}$$

where the label harmonized according to Fig. 2. The authors have concluded that a much reduced and more simple version of the P6 gives slightly worse results than the P13 version. DARC is adaptive predictor which adjusts the prediction based on a simple estimation of horizontal and vertical gradient [2]:

$$\begin{aligned} g_v &= |W-NW|, \\ g_h &= |N-NW|, \end{aligned} \tag{5}$$

and the corresponding weighted coefficient $\alpha = g_v/(g_v+g_h)$. The prediction is then performed by applying the equation:

$$P=\alpha W+(1-\alpha)N\,. \tag{6}$$

Predictors based on Minimum Mean Square Error – MMSE perform the adaptation prediction coefficient on the basis of a training set of causal pixels. This approach can achieve better results for entropy compared to the using a fixed number sub predictors. However, in adaptation, is necessary to count pseudo inversion whose order increases with increasing training set. If we use last m pixels for the adaptation of predictor coefficients k-th order, vector of coefficients is calculated according to the expression:

$$\boldsymbol{a}=\left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\left(\boldsymbol{C}^T\boldsymbol{y}\right), \tag{7}$$

where the matrix of training set order is $m\times k$ marked as $\boldsymbol{C}$, and vector last m pixels value indicated by $\boldsymbol{y}$.

In the case of compressing 3D medical images, in [6] simple solution, based on the differential code modulation, which exploits the spatial redundancy in all three dimensions is proposed. Algorithm called 3DPCM made the prediction in all three coordinates based only on previous pixel value. For each frame of the 3D picture, firstly, prediction is done with pixel W for species, then with pixel N for columns. Redundancy between frames is reduced by the prediction of each pixel based on the value of that pixel from the previous frame.

Here we have proposed a solution that takes advantage of the MED and GAP predictors, i.e. simplicity of the median predictor and advantages of gradient estimation in GAP predictor. As MED predictor, chooses between the context of vertical edges, horizontal edges and smooth regions. Selection is done by simple estimation of horizontal and vertical gradient and one threshold. The number of contextual pixel is also a compromise between the MED and GAP predictors, and in contrast to the GAP predictor, which has three predefined threshold, the proposed predictor is based on one threshold. Gradient estimation is based on the equation, predictor is based on edge detection, so we adopt the name of the Gradient Edge Detection (GED). GED local gradient estimation is done using following equations:

$$\begin{aligned} g_v &= |NW - W| + |NN - N|, \\ g_h &= |WW - W| + |NW - N|. \end{aligned} \tag{8}$$

Local gradient estimation is followed by a simple predictor. The prediction is done using the algorithm:

$$\begin{aligned} &if\ g_v - g_h > T, P = W \\ &elseif\ g_v - g_h < -T, P = N \\ &else\ P = 3\frac{N+W}{8} + \frac{WW+NW+NN}{12}. \end{aligned} \tag{9}$$

where T is the designated threshold, which can be predefined or may be ordered based on the context. Second variant, named GED2 predictor, is slightly different from the first one in the case when smooth region is detected. For the estimation of local gradient, is also used equation (8). Complicated and inefficient equation given in (9) is replaced by same equation that uses MED predictor in the case of smooth region. Therefore, GED2 prediction is made by following equation:

$$\begin{aligned} &if\ g_v - g_h > T, P = W \\ &elseif\ g_v - g_h < -T, P = N \\ &else\ P = N + W - NW. \end{aligned} \tag{10}$$

GED2 predictor is a simple combination of gradient and median predictors. Estimation of local gradient and a threshold is used to decide which of the three sub predictors is optimal, i.e. is the pixel in context of horizontal edge, vertical edge or smooth region.

4 Comparative Analysis

Lossless image compression has to preserve the exact value of each pixel, regardless of whether there is noise or not. Measure performance predictor can be expressed over the degree of compression, i.e. relations between required memory space before and after compression, or equivalently by bit rate, which is the average number of bits needed to code a single pixel. However, predictors evaluation with these parameters was not possible to use. Predictor only eliminates redundancy, and in fact does not do compression. As a measure of predictor efficiency entropy of prediction error image is used. Assuming that the image can be modeled as Markov model without memory, entropy is the minimum possible bit rate that can be achieved after the coding. If the picture denote as random variable $\boldsymbol{X}$ with alphabet $A=(a_0,a_1,a_2,\ldots,a_{N-1})$, which means that we're processing of N-bit image, we can calculate the entropy:

$$H(X)=-\sum_{x\in A}p(x)\log_2 p(x), \tag{11}$$

where $p(x)$ indicates the probability of occurrence of symbol x, i.e. level of grayness value x, in our case.

Prediction algorithms described in the previous chapter were tested on a set of uncompressed medical images. In practice, medical images can be a series of 2D data, i.e. slices. Those often use as sequential slices with low spatial step, and then they make 3D image, Fig. 3. Therefore, series images of computer tomography (CT) and magnetic resonance imaging (MRI) are used for testing. As an output parameter, the total entropy of prediction error 3D image has calculated. Test images set contains a 150 previously uncompressed medical images, some of which are shown of Fig. 4.

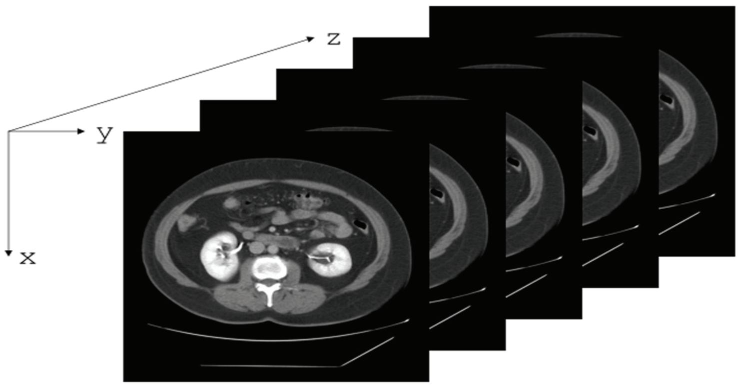


Fig. 3 – *3D medical image.*

Medical images are usually represented as a 12-bit or 16-bit images, and gradient adaptive predictor used fixed thresholds defined with 8-bit images. Therefore the version with the scaling gradient estimation is tested, as described in [4]. If N-bit image is compressed, the authors in [4] have suggested scaling the gradient estimation based on the empirical relation:

$$\lambda=2^{-\left\lfloor\frac{N-8}{2}\right\rfloor-\max\left(0,\lceil\log_2\sigma-5\rceil\right)}, \tag{12}$$

where σ is a total absolute prediction error prior scanned row. **Table 1** provides an overview of performance of different predictors applied to 3D medical images described with entropy. In addition to MED predictor, two versions of GAP predictor are tested, the basic version described by (2) and (3), and version GAPs scaled with (12). The proposed GED predictor is tested with different thresholds, which are selected to be degree of number 2. Mark GED_{16} says that it is GED predictor with threshold 16. Similarly, the second alternative predictor is marked, e.g. $GED2_{16}$ is GED2 predictor with threshold 16. Also, described predictors DARC, P6 and DPCM3d are tested, and predictors based on the MMSE method are not considered as damaging the principle of simplicity.

By analysis of the results in **Table 1** can be concluded that the best performance has predictor P6 based on the penalty and the combination of several simple sub predictors. However, this predictor is much more complicated than the other examined predictors. Scaling method of gradient estimation GAP predictor, given by (12), does not guarantee efficient prediction, although considerably complicates prediction algorithm. Also, is noticeable that simple predictors, such as those in the MED, P6 and GED2, give a better prediction than more complicated predictors, such as those in the GAP or GED.

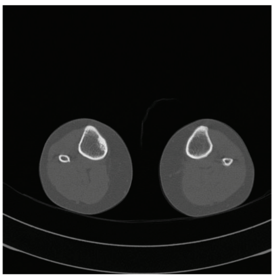
Table 1
Comparative analysis entropy of prediction error
3D *medical images, applying different predictors.*

Predictor	CT1	CT2	CT3	MRI1	MRI2
MED	7,615	5,796	3,998	4,747	4,253
GAP	7,563	5,727	4,304	4,851	4,375
GAPs	7,507	5,705	4,305	4,976	4,514
GED16	7,606	5,779	4,598	5,072	4,684
GED64	7,495	5,758	4,711	5,264	4,876
GED128	7,445	5,787	4,786	5,359	4,945
GED216	7,754	6,007	4,150	4,847	4,278
GED264	7,893	6,032	3,918	4,738	4,170
GED2128	7,926	6,025	3,788	4,702	4,144
DARC	7,725	5,937	4,522	5,044	4,596
P6	7,532	5,703	3,706	4,643	4,107
DPCM3d	7,755	6,322	5,558	6,238	5,745

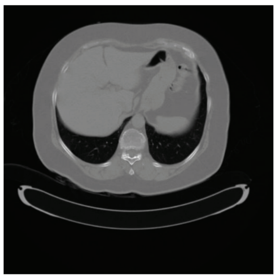
Table 2
Entropy mean value after prediction of 150 CT *images with different predictors.*

Predictor	MED	GAP	GAPs	GED16	GED64
Entropy	3,953	4,199	4,210	4,475	4,606
Predictor	GED216	GED264	GED2128	DARC	P6
Entropy	4,078	3,883	3,780	4,417	3,690

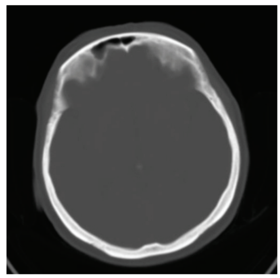
Table 2 provides the analysis of predictors performances in a large number of independent 2D medical images. Set of 150 2D CT images are tested, and the final entropy is the result of averaging individual image entropy. Examples of used medical images are shown on Fig. 4.



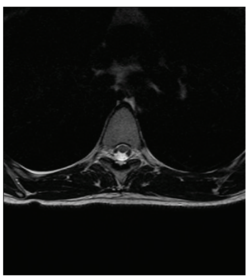
CT1 512x512x38



CT2 512x512x50



CT3 512x512x14



MRI1 640x576x25

MRI2 378x384x15

Fig. 4 – *Examples of medical images.*

5 Conclusion

This paper describes predictive lossless image compression process. Also, the most important predictors of the most important algorithms for lossless compression are described, which are accepted in standards or that are representative to their characteristics. Novel solution for the simple linear prediction is based on the detection of edges, called the GED and its comparison with the described predictors is made. GED algorithm is a mixture of distinguish features of most representative linear predictors, namely MED and GAP. Proposed predictor has shown satisfactory results on a chosen set of uncompressed medical images.

6 Acknowledgment

This paper is the result of research during the project called Lossless image compression, which is supported by the Ministry of Science and Technology of the Republic of Srpska, No. contract 06/0-020/961-166/09. Also, the authors wish to thank KBC Paprikovac, Banja Luka, who supplied medical images for testing that are used in this research and professor Branimir Reljin, School of Electrical Engineering in Belgrade, on his useful suggestions.

7 References

[1] I.N. Bankman: Handbook of Medical Image Processing and Analysis, Elsevier, London, 2009.

[2] N. Memon, X. Wu: Recent Developments in Context-based Predictive Techniques for Lossless Image Compression, The Computer Journal, Vol. 40, No. 2/3, 1997, pp. 127 – 136.

[3] M.J. Wienberger, G. Seroussi, G. Sapiro: LOCO-I: A Low Complexity, Context-based, Lossless Image Compression Algorithm, Conference on Data Compression, Snowbird, UT, USA, March/April 1996, pp.140 – 149.

[4] X. Wu, N. Memon: CALIC – A Context based Adaptive Lossless Image Codec, IEEE International Conference on Acoustics, Speech, and Signal Processing, Atlanta, GA, USA, Vol. 4, May 1996, pp. 1890 – 1893.

[5] G. Deng, H. Ye: Lossless Image Compression using Adaptive Predictor Combination, Symbol Mapping and Context Filtering, Conference on Image Processing, Kobe, Japan, Vol. 4, Oct. 1999, pp. 63 – 67.

[6] V. Nandedkar, S.V.B. Kumar, S. Mukhopadhyay: Lossless Volumetric Medical Image Compression with Progressive Multiplanar Reformatting using 3D DPCM, National Conference on Image Processing, Bangalore, India, March 2005.