CTA'07, April 12-14, Hammamet, Tunisia

Lossless and Lossy Compression in 4D Bio-modeling

Martin Žagar, Mario Kovač, Ivana Bosnić
Department of Control and Computer Engineering in Automation
Faculty of Electrical Engineering and Computing
Unska 3, HR-10000 Zagreb, Croatia
martin.zagar@fer.hr

Abstract

Bio-modeling has important role in virtual reality in telemedicine providing base for simulations and decisions. Virtual reality is based on sequences of volumetric images whose motion is captured in time. These data sets are typically very large in size and demand a great amount of resources for storage and transmission. Therefore it is necessary to compress this data both fast and efficiently. We will propose combination of lossless and lossy compression models to obtain toset demands.

Key-words: 4D bio-modeling, lossless and lossy compression, determination area of interest

1. Introduction

Bio-modeling is the use of computer modeling programs for biological applications. It is based on digital three-dimensional (3D) images of biological structures that can be rendered, viewed, manipulated, and physically reproduced. These 3D renderings may be created using a computer tomography (CT) scanning, electron beam tomography (EBT) scanning, magnetic resonance imaging (MRI), electron beam tomography, hand-held or stationary 3D laser scanners. If we add time-scale and observe changing of 3D structures through time, we get 4D bio-modeling. Four dimensional (4D) medical data are sequences of volumetric images whose motion is captured in time. These data sets are typically very large in size and demand a great amount of resources for storage and transmission. 4D biomedical data describe temporal changes of structures or processes (such as functional activation areas) as a sequence of three-dimensional (3D) images or volumes. Similarly, 3D medical images are a sequence of two dimensional (2D) image slices that represent cross sections of a structure. Without efficient compression, large amounts of data due to sequences of 4D medical images, would easily overwhelm storage and transmission systems. For efficient storage and transmission of such data, compression algorithms are imperative. Various compression methods are available in the literature. We can group them in 2 classes: lossy and lossless. Lossy compression algorithms characterize different retrieved data while decompressing, from original. Lossy algorithms can be applied on coding data of smaller interest. Lossy algorithms perform better than lossless ones in terms of compression efficiency, but can not be used in simulations where original quality must be retained. Allowing for a user controlled error in compressing, we can achieve a dramatic, data dependent, decrease in the required storage space with a controlled impact in the image quality. Lossless compression schemes are used when exact original data is needed to be reconstructed from the compressed data. Lossless algorithms provide reliability in coding of vital points of interest and therefore can be used for compressing important areas of medical 3D data enhanced with fourth dimension – time. Lossless data compression schemes often consist of two distinct and independent components: modeling and coding. 3D motion estimation, as a preprocessing step, effectively removes the redundancy between frames and so reduces the corresponding prediction error. Various coding techniques can be applied on residual images produced by 3D motion estimation. However, most volumetric datasets are used in biomedicine and other scientific applications where lossy compression is unacceptable. In these cases, lossless compression algorithms must be used.

2. Goals of compression in bio-modeling

We assume that application data is stored as one or multiple arrays, i.e. temporal series. The elements of the arrays are values that change like polynomial function. These arrays contain a sequence of values of this function and are typical for numerical dynamic simulations of physical phenomena where scalar values change in time.

Elements of an array are sometimes computed from respective elements of another array so that the correlation between the values can be revealed. These numerical properties might result in very high compression ratio. The problem is how to reveal all the features of the data and how to use them for data compression.

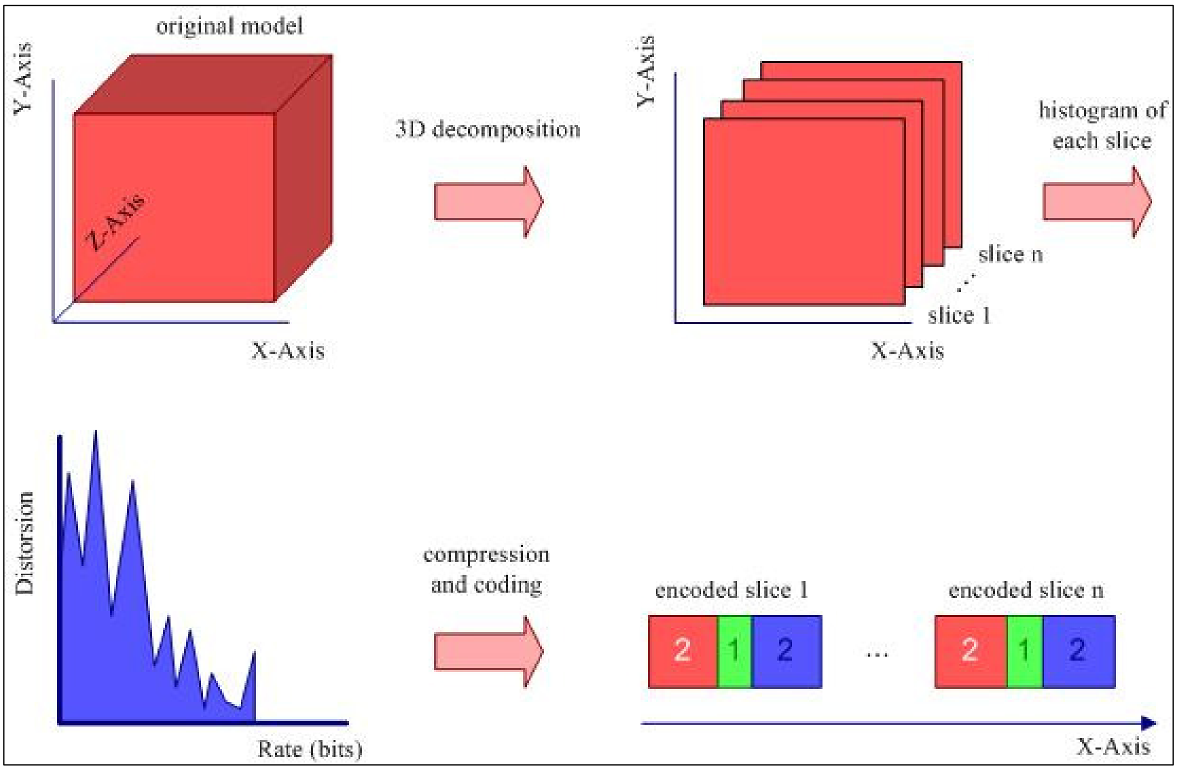

Figure 1. Compression schema for encoding original model into 1-D code

3. Compression schema

4D medical images may also be treated as video sequences, with each image slice of each volume representing a frame and the fourth dimension representing time. Motion compensation is an efficient way to reduce the temporal redundancy among different frames in a video sequence. The objective of the motion compensation process is to determine the amount of motion on a block by block basis which minimizes the difference between consecutive frames. First, a picture is split into blocks (Figure 1.). The first picture of a sequence is usually encoded as an meta frame using only information contained in the picture itself. The residual information resulting from motion compensation or spatial prediction is transformed to frequency domain and the resulting coefficients are quantized. Variable length coding techniques are used to perform entropy coded. Typical 4D datasets exhibit a high degree of temporal coherence. The rendering application can be accelerated if only the subvolumes that actually change over time are located and rendered. We propose a method to achieve that, using slices. The slices cache the already rendered unchanged regions of the 4D dataset. Changed subvolumes are updated to the volume and subsequently rendered. Finally the slices are recomposited to produce the final image. The compression algorithm works as follows. We can use the fact that the difference between the array elements is small relatively to the element values. For each volume, an optimal linear predictor (in the mean-square error sense) is determined by solving a linear system of equations featuring estimated covariance values of the volume. This optimal linear predictor is used to calculate a sequence of difference values which are then encoded using H.264/AVC code, producing the compressed output.

H.264/AVC method of compression can be adopted on two encoding methods for lossless compression of 4D medical images. Both methods take

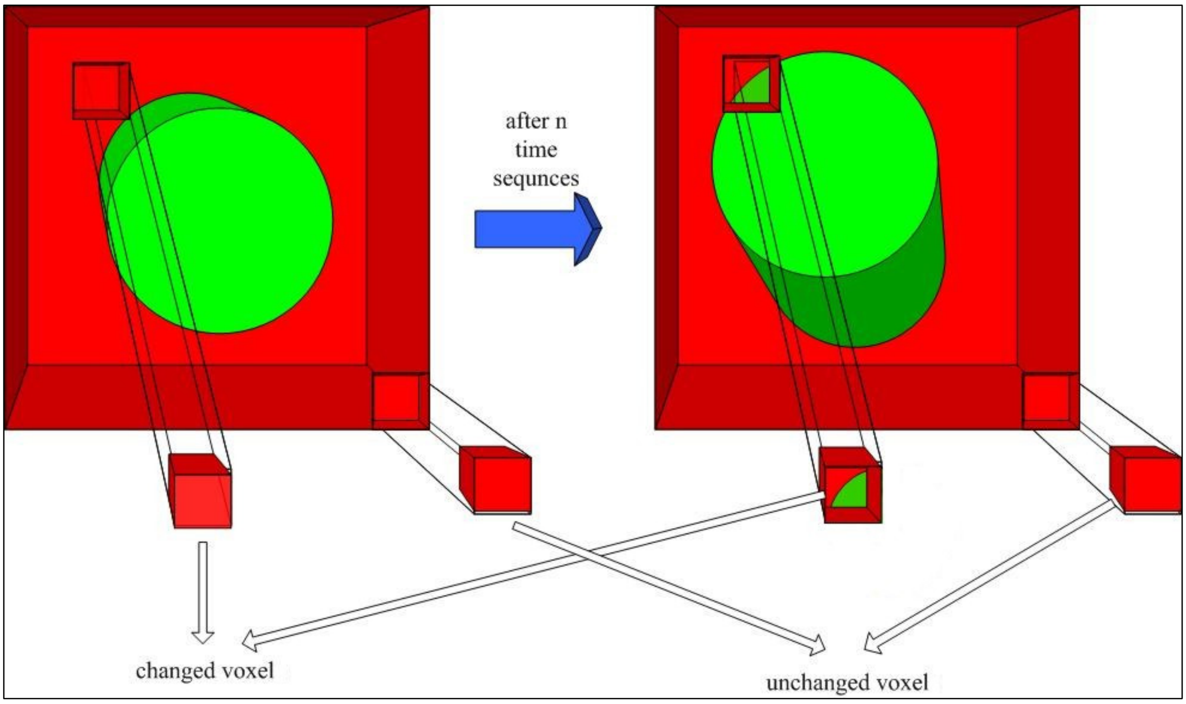

Figure 2. Defining area of high interest

advantage of the motion compensation model implemented in H.264/AVC. The first method, referred to as H.264-VOL hereafter, exploits the redundancy among image slices within each volume. The second method, referred to as H.264-TIME hereafter, exploits the similarities between image slices in the fourth dimension (time). Method H.264-VOL exploits the redundancy among image slices within volumes of 4D medical images by applying H.264 along the third dimension. Each volume is encoded separately as a complete video stream and it may be decoded without having to decode the whole 4D data set [2]. On the other hand, method H.264-TIME considers the temporal dimension of the 4D medical image. H.264-TIME applies H.264 across the fourth dimension (time) by grouping together all similar image slices across all volumes. Thus, the first group or “video stream” starts with the first image slice of the first volume, followed by the first image slice of the second volume and so on, and ends with the first image slice of the last volume. Therefore, each group of similar image slices is encoded separately and it may be decoded without having to decode the whole 4D data set [2].

Since a 4D medical image describes the dynamic changes of a 3D image in time, the redundancies across the fourth dimension are higher than within 3D images. Algorithms for 3D coding (such as JPEG2000) are trying to find redundancy in only 3 dimensions, but if want to achieve the highest compression rate, we must find redundancy in all 4 dimensions, including time, i.e. temporal dimension. Object-based processing concerns both transformation and coding. In the perspective of transformation it brings up a boundary problem. As discrete signals are nothing but sets of samples, it is straightforward to associate the idea of object to a subset of samples, usually sharing some common features. Therefore we propose use of direct predecessor and successor values to determine boundaries and coding schema. Applied on H.264-TIME, it is expected to yield a better compression performance than the 3D case coding. A typical transform coding algorithm consists of three major stages: transform, quantization, and encoding. An input dataset is passed through some transformation to represent it using a different mathematical basis. This new representation will reveal the correlation that exists in the data. The coefficients of correlation produced in this stage, are quantized to produce a stream of symbols, each of which corresponds to an index of a particular quantization bin [4]. The last stage encodes the stream of symbols and attempts to represent it as efficiently as possible, both in lossless and lossy way of compression (Figure 1.).

4. Defining objects of interest

Object-based processing concerns both transformation and coding. In the perspective of transformation it brings up a boundary problem. As discrete signals are nothing but sets of samples, it is straightforward to associate the idea of object to a subset of samples, usually sharing some common features. Therefore we propose the use of direct predecessor and successor values to determine boundaries which determine the area of interest and according coding schema.

In a set of head MRI images for example, the object of interest is most probably the brain, and the rest of the image can be considered as the background. If the “average” model for the generic patient were available to both the encoder and the decoder, only the deformation parameters needed to fit it to the current data would need to be transmitted. In this framework, the weight assigned to a voxel depends on its semantics [6].

The impact of the nonrelevant information must not be underestimated. This is particularly important when dealing with medical images, where the background often encloses the majority of the voxels. For a typical MRI dataset for instance, about the 90% of the voxels belong to the background (Figure 2.). It is thus of prime importance to classify them a priori in order to assign higher interest priority. Coding efficiency results from the tradeoff between the improvement due to the separation of sources with different statistics and the degradation due to the overhead implied by the border voxels.

Medical images usually consists of a region representing the part of the body under investigation (i.e., the heart in a CT or MRI chest scan, the brain in a head scan) on an often noisy background with no diagnostic interest. It seems thus very natural to process such data in the object-based framework: assign high priority to the semantically relevant object, to be represented with up-to-lossless quality, and lower priority to the background.

In the proposed system, the object of interest and the background are encoded independently. Each generates a self-contained segment of the bitstream. This implies that the border information is encoded twice: as side information for both the object and the background. In this way, each of them can be accessed and reconstructed as if the whole set of wavelet coefficients were available, avoiding artifacts along the contours for any quantization of the decoded coefficients. The encoding time depends on the context, and increases with the size of the neighborhood. Efficiency can thus be improved by choosing spatial conditioning terms of small support.

Basically, the subband coefficients within the area of interest mask are shifted up (or, equivalently, those outside this area are shifted down) so that the minimum value in the area of high interest is greater than the maximum value in the background. This splits the bitplanes respectively used for the area of interest and the background in two disjoint sets. The rate allocation procedure assigns to each layer of each codeblock (in the different subbands) a coding priority which depends on both the semantics and the gain in terms of rate/distortion ratio. This establishes the relative order of encoding of the high interest area subband coefficients with respect to the background.

For example, for the head MRI dataset high priority is assigned to the background layers in the codeblocks, moving the focus of the encoder out of the area of interest. That area and background codeblocks are mixed up, compromising high important-based functionalities. This can be easily verified by decoding the portion of the bitstream indicated by the encoder as representing the area of interest. The resulting image is composed of both the high important area and the background.

Very important thing in lossy compression is rate distortion function. Classical rate-distortion theory evolved from Shannon’s theory of communication. It studies the tradeoff between the rate and the achievable fidelity of the transmitted representation under some distortion function, where the analysis is carried out in expectation under some source distribution.

If that rate is low, we may still be able to transmit objects that have a very regular structure without introducing any distortion, but this becomes impossible for objects with high information density. In applications of lossy compression, one may be willing to accept a rate which is lower than the minimal sufficient statistic complexity, thereby losing some structural information. However, for a minimal sufficient statistic *y*, theory does tell us that it is not worthwhile to set the rate to a higher value than the complexity of *y*.

The shape of the codelength function for an object *x* is more complicated. In theory, the codelength can never become less than the complexity of *y*, and the minimal sufficient statistic witnesses the codelength function at the lowest rate at which the codelength is equal to the complexity of *y*.

It can be seen that a temporary increase in the codelength function can occur up to a number of bits that depends on the so-called covering coefficient. Loosely speaking this is the density of small distortion balls that is required in order to completely cover a larger distortion ball. The covering coefficient in turn depends on the used distortion function and the number of dimensions.
Adaptable compression schema based on defining regions of interest and lossless coding of the high interest regions and lossy coding of background, compared with regular lossless coding is shown in the Table 1.

Table 1. Comparison of used bits per pixel in coding of test pictures

Picture	Lossless coding	Adaptable compression schema
MR-picture1	5.73	5.29
MR-picture2	5.25	4.86
CT-picture	6.34	5.96

5. Examples of usage

Bio-modeling is very important part of telemedicine. In implementation of telemedicine system, making the appropriate models, their coding and transmission to centre where expert knowledge is located, are crucial for improving quality of service.

Bio-modeling incorporates three operations for managing data sets, creating models and analyzing results: data management, model fitting and expert analysis. In each step compression algorithms enable faster communication between operations and enable more efficient usage of the storage devices. Examples come from clinical and preclinical trials where biomedical data display subject and residual unknown variations.

Heterogeneity of platforms, problems, interfaces causes difficulties in visualization of data and models but coding with compression algorithms is same on each platform and the most important thing, which is decoding to original (or pseudo-original image for backgrounds), is also platform independent.

Four dimensional space-time volume data appears also frequently in computational fluid dynamics and global climate simulations. Sampled light fields or lumigraphs created for image-based rendering are also volume data in 4D. Interactively handling and visualizing such datasets has become increasingly important.

4D bio-modeling can be used in various health systems, analysis of mutations in cancer, prediction of protein structure, gene expression, measuring biodiversity and bioinformatics and computational biology.

It is obvious that telemedicine might be helpful in monitoring the early postoperative course of patients discharged to their home care. By transmitting of temporal bio-models from the outpatient facility to the hospital, and by providing consultations between the physicians and the specialists we were be able to give proper and prompt medical treatment without moving patients from the remote areas. Finally bio-modeling and telemedicine, according to cost-benefit analysis is very cost-effective.

Remote consultations from a rural clinic to a specialist can alleviate prohibitive travel and associated cists for patients. Bio-modeling can also opens up new possibilities for continuing education or training of isolated or rural health practitioners, who may not be able to leave a rural practice to take part in some professional meetings or in educational courses.

Full-body motion interfaces are used to simulate movement of the whole body. In that cases it is relevant to decide which areas have high importance and to distinguish objects from background for successful compression and coding the whole body.

Motion systems provide passive user movement through the virtual environment. Some characteristic examples of such applications are virtual endoscopy, computer-aided surgery, dentistry, radiotherapy planning, rehabilitation and other levels of telemedicine.

Visible human data have been used in many projects as a test data set. Temporal 3D visualizations of various anatomical parts have been used for education of medical students. Various organ models have been developed using this data.

6. Conclusion

Realization of 4D bio-modeling is based on combination of three main enabling technologies. First of all, fast computer and communication networks, then human-computer interface which is especially important in modeling step (producing models) and applications of algorithms that are applied on these models enable storage and communication.
Coding of 4D bio-models should be both fast and efficient. In this paper is shown that with defining the regions of interest there can be achieved higher compression rates. Therefore we proposed lossless coding of areas of high interest and lossy coding areas of small interest (background). The interest-based processing enables a pseudo-lossless regime where the object of interest can be encoded/decoded without loss, and combined with the background that can be represented at a lower quality.

In this way, images respecting the lossless constraint in the high interest areas and preserving a good visual appearance in the background can be obtained at a significantly lower rate. Each object is encoded independently to generate a self-contained segment of the bitstream.

It is highly recommended to select a compression technique that additionally provides a multi-resolution representation. This offers the basis for processing based on level of interest, and accordingly details, of compressed data.

7. References

[1] J. E. Fowlert, R. Yagel, "Optimal Linear Prediction for the Lossless Compression of Volume Data", *Data Compression Conference, DCC '95* Proceedings, Snowbird, UT, USA, 1995., pp. 458

[2] V. Sanchez, P. Nasiopoulos, R. Abugharbieh, "Lossless Compresion of 4D Medical Images Using H.264/AVC", *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings*, Toulouse, France, 2006., pp. II-1116-II-1119

[3] L. Ibarria, P. Lindstrom, J. Rossignac, A. Szymczak, "Out-of-core compression and decompression of large n-dimensional scalar fields", *Eurographics 2003,* Granada, Spain*,* 2003.

[4] S. de Rooij, P. Vitanyi, "Approximating Rate-Distortion Graphs of Individual Data: Experiments in Lossy Compression and Denoising", *Computer Science*, abstract cs.IT/0609121, 2006.

[5] P. Yan, A. Kassim, "Lossless and Near-Lossless Motion-Compensated 4D Medical Image Compression", *2004 IEEE International Workshop on Biomedical Circuits and Systems,* 2004.

[6] G. Menegaz, J. P. Thiran, "Lossy to Lossless Object-Based Coding of 3-D MRI Data", *IEEE Transactions on Image Processing*, vol. 11, num. 9, 2002, pp. 1053-1061

[7 L. Zeng, C. Jansen, M. A. Unser, P. Hunziker, "Extension of wavelet compression algorithms to 3D and 4D image data: exploitation of data coherence in higher dimensions allows very high compression ratios", *Proc. SPIE* Vol. 4478, pp. 427-433, Wavelets: Applications in Signal and Image Processing IX, 2001.

[8] C. Bajaj, I. Ihm, S. Park, "3D RGB Image Compression for Interactive Applications", *ACM Transactions on Graphics (TOG),*Volume 20, Issue 1, 2001.

[9] K. Anagnostou, T. J. Atherton, A. E. Waterfall, "4D Volume Rendering With The Shear Warp Factorisation: Extensions And Quantitative Results", *Fifth International Conference on Information Visualisation,*. Proceedings. London, 2001.

[10] J. E. Fowler, R. Yagel, "Lossless compression of volume data", *Symposium on Volume Visualization*, Proceedings of the 1994 symposium on Volume visualization, pp. 43-50, Tysons Corner, Virginia, USA

[11] V. Engelson, D. Fritzson, P. Fritzson, "Lossless Compression of High-volume Numerical Data from Simulations", *Data Compression Conference*, DCC 2000, pp. 574-587, Proceedings, Snowbird, Utah, USA, 2000.

[12] C. Harless, J. J. DiStefano, "Automated expert multiexponential biomodeling interactively over the Internet", *Computer Methods and Programs in Biomedicine*, Volume 79, Issue 2,pp. 169-178, 2005.

[13] O. Hadar, A. Stern, R. Koresh, "Enhancement of an image compression algorithm by pre- and post-filtering", *Optical Engineering,* Volume 40, Issue 2, pp. 193-199, 2001.

[14] S. Loncaric, "Virtual Reality in Medicine", *Telemedicine*, pp. 455-475, Telemedicine Association, Zagreb, Croatia, 2005.