An optimized 3D context model for JPEG2000 Part 10

T. Bruylants[*a], A. Alecu[a], T. Kimpe[b], R. Deklerck[a], A. Munteanu[a], P. Schelkens[a]
[a]Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
[b]Barco, Pres. Kennedypark 35, 8500 Kortrijk, Belgium,

ABSTRACT

The JPEG2000 standard is currently widely adopted in medical and volumetric data compression. In this respect, a 3D extension (JPEG2000 Part 10 – JP3D) is currently being standardized. However, no suitable 3D context model is yet available within the standard, such that the context-based arithmetic entropy coder of JP3D still uses the 2D context model of JPEG2000 Part 1. In this paper, we propose a context design algorithm that, based on a training set, generates an optimized 3D context model, while avoiding an exhaustive search and at the same time keeping the space and time complexities well within the limits of today hardware. The algorithm comes as a solution for the situations in which the number of allowable initial contexts is very large. In this sense, the three-dimensional 3x3x3 context neighborhood investigated in this paper is a good example of an instantiation that would have otherwise been computationally unfeasible. Furthermore, we have designed a new 3D context model for JP3D. We show that the JP3D codec equipped with this model consistently outperforms its 2D context model counterpart, for an extended test dataset. In this respect, we report a gain in lossless compression performance of up to 10%. Moreover, for a large range of bitrates, we always obtain gains in PSNR, sometimes even over 3dB.

Keywords: compression, multi-resolution and wavelets, mutual information, JPEG2000, volumetric, context modeling

1 INTRODUCTION

The medical and scientific imaging equipment available today generate a considerable amount of data that needs to be stored, transmitted and accessed. It is clear then that an efficient representation technology is required to allow for optimal storage or transmission, efficient random access, region-of-interest (ROI) support and resolution/quality scalability. JPEG2000 is an image coding system that uses state-of-the-art compression techniques based on wavelet technology. Its architecture lends itself to a wide range of uses, including medical imaging. With the approval of JPEG2000 as an accepted image compression option by the DICOM WG 4, it is now available to an increasingly wider medical user community. The JPEG2000 specification is divided into different parts, with Part 1 being the 2D core codec [1]. Currently, JPEG2000 provides the required functionality for two-dimensional data sets through Part 1 and Part 2 [2]. In addition, even though Part 2 gives indirect support for a three-dimensional wavelet transform by use of the arbitrary component transformation, a complete volumetric coding scheme is still essential to provide all the required functionality, as well as an optimal rate-distortion performance and an isotropic behavior. With Part 10 of JPEG2000 [3] an effort was taken to provide the same functionality and efficiency for three-dimensional data sets. Part 10 (JP3D) is the extension of JPEG2000 that will provide isotropic support for rectangular three-dimensional data sets with no time component. The JP3D specification is currently in the Final Committee Draft (FCD) stage, meaning that all features are defined and finalized. One pitfall, however, is that the current JP3D standard still uses the 2D context model of Part 1 to provide contexts to the arithmetic coder. This 2D context model is applied to the XY planes of the code-blocks being entropy coded. A new improved context model is not yet included, because no such suitable context model exists yet. Nonetheless, previous research [4] shows that a significant gain in compression performance is still possible with a context model that uses state information from all three dimensions. If such a context model is found, it will be added to the JPEG2000 Part 10 specification as an addendum. This paper reports about the ongoing research on the topic of three-dimensional context modeling. In this respect, we propose (i) a new context design algorithm and (ii) a new 3D context model for JP3D. The former differs from existing methods by the fact that, unlike its predecessors, it supports context models for higher dimensions, while keeping complexity and execution time within the limits of today's hardware.

[*] tim.bruylants@vub.ac.be

Additionally, we show that the 3D context model designed using this algorithm significantly improves the compression performance.

The paper is organized as follows: section 2 gives an overview of the JPEG2000 standard and the context models described within; we discuss the proposed context design algorithm in section 3; experimental results are illustrated in section 4; finally, we give our conclusions in section 5.

2 JPEG2000 CONTEXT MODELING

2.1 Embedded Block Coding by Optimized Truncation (EBCOT)

The JPEG2000 compression scheme consists of a discrete wavelet transformation followed by quantization. The samples of the resulting subbands are then divided into code blocks. The EBCOT coder is then used so as to independently encode each of these code blocks. EBCOT is a two-tiered context-based adaptive arithmetic coder, in that tier-1 contains the actual context modeling and bit coding, while tier-2 handles the data ordering and truncation. For the scope of this paper, we will only consider tier-1. Thus, let us first note that each code block is independently fractionally bitplane coded, starting from the most significant bitplane with a non-zero element to the least significant bitplane. Each coefficient bit in a bitplane is scanned in a stripe-based order and is coded by only one of three coding passes. These passes make use of four coding primitives to actually encode the bits, namely the zero coding (ZC), the sign coding (SC), the magnitude refinement coding (MR) and the run-length coding (RLC). Each primitive delivers one or more pairs of a binary symbol and a context label, respectively, to the arithmetic coder. We refer the interested reader to literature [5] for an in-depth discussion of the EBCOT tier-1 architecture.

2.2 3D Context models

Recall that the JPEG2000 Part 1 specification describes a 2D context model. It has been shown in [6] that for two dimensions, the preferred wavelet neighborhood is given by the set of 3x3 coefficients located around the central sample. Hence, for the ZC primitive, this model uses the significance state of the eight neighbors in the XY plane surrounding the samples to generate context labels. The significance state of a sample is a binary value that initially has the value of insignificant and changes to significant as soon as the first significant bit of the sample is coded [5]. In [4] it is shown that, when making use of 3D context modeling for the ZC primitive in an EBCOT codec, compression gains of up to 10% can be achieved. Nonetheless, the 3D context model used in [4] is a heuristic, straightforward extension of the standard 2D context model used by JPEG2000. Figure 1 shows the 3x3x3 context neighborhood with the neighbors classified in six groups. The X, Y and Z labels group the two neighbors that lie on the X-, Y and Z-axis respectively. The XZ, YZ and XY labels represent the four neighbors that lie in the XY, YZ and XY planes respectively. Finally, the group labeled as XYZ contains the eight corner neighbors of the 3x3x3 cube. The six groups indicated in Figure 1 are used to formulate the selection criteria on which the 3D context model of [4] is built. Tests carried out with JP3D equipped with this model, reveal that the latter does not perform well. The explanation for this consists in the fact that the EBCOT algorithm used in [4] employs an additional Quad Tree Splitting (QTS) step, in combination with which the proposed context model behaves well. However, as QTS techniques finally have not been adopted within the JPEG2000 standard, it is clear that an alternative 3D context model should be designed for JP3D. In this respect, it should be first pointed out that significant compression gain can only be achieved for the ZC primitive, as the use of the SC, MR and RLC primitives is not dependent on the actual number of dimensions in the data. Moreover, the main issue for finding a working 3D context model is that the required space and time complexity exceeds modern machine capabilities, when performing an exhaustive search.

This paper describes a formal mechanism to generate an optimized 3D context model for the ZC primitive, while additionally keeping complexities low. The obtained classification scheme is a large table that assigns all possible contexts, I, to F context labels.

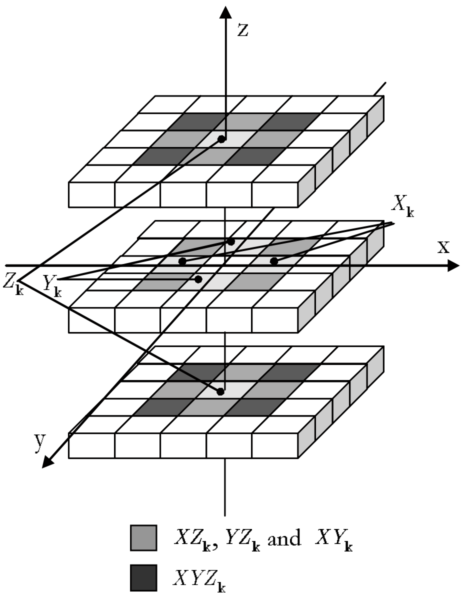

Figure 1. A three-dimensional 26-bit preferred neighborhood.

3 A LOW-COMPLEXITY NEAR-OPTIMAL CONTEXT MODELING ALGORITHM

3.1 Context modeling based on Mutual Information

The 3D ZC context model formation presented in this paper is based in part on the work of [7], in which they analyze the standardized 2D context model of JPEG2000, by comparing it with their newly constructed 2D context model. Thus, the methodology we employ uses the concept of maximizing the mutual information (MI) estimates [8] between possible context classifications. Starting with an initial number of contexts I and the desired number of context groups F, it can be shown that the total number of possible classification schemes is given by $S(I,F)$ where S is the Stirling number of the second kind. The optimal classification scheme is then a trade-off between having maximal MI and a minimal F. However, it should be pointed out that for *n*-dimensional context models, where $n > 2$, the number of initial contexts I explodes. Indeed, in 2D, the ZC primitive uses the significance status of the eight neighbors to form a context group template. As stated before, each neighbor's significance status can take two values: *significant* or *insignificant*. The 2D work of [7] proposes a bottom-up approach that starts with a context classification of $I = 256$ context groups, which are then sorted according to the conditional probability $P(1|c_i)$, i.e. the probability that a "1" symbol is encoded within context c_i. In order to calculate the conditional probability $P(1|c_i)$, a Joint Probability Matrix (JPM) is first constructed, using a representative data training set, that contains the probabilities $p(0,c_i)$ and $p(1,c_i)$, for all possible contexts (i.e. $0 \le i \le I$). Thus the JPM is a $2 \times I$ matrix of probability values. After that, the classification is refined by joining in-order-adjacent context groups while keeping the mutual information maximal. The joining process is repeated until the requested amount of context groups is reached.

It can be easily proven that a bottom-up approach toward finding an optimized classification will lead to a total number of searches given by

$$\sum_{x=F}^{I} x = \frac{I(I+1)-(F-1)F}{2} \tag{1}$$

Hence, for $I = 256$, the number of searches (1) still remains within reasonable computational limits. However, it can be observed that for 3D, when using a 3x3x3 context neighborhood, as shown in Figure 1, $I = 2^{26}$ different context values can be devised. It is clear then that for large values of I, a bottom-up approach is no longer a viable computational solution. As such, a concern of this paper is to devise an algorithm that allows the design of optimized context models, applicable for a large number of initial contexts.

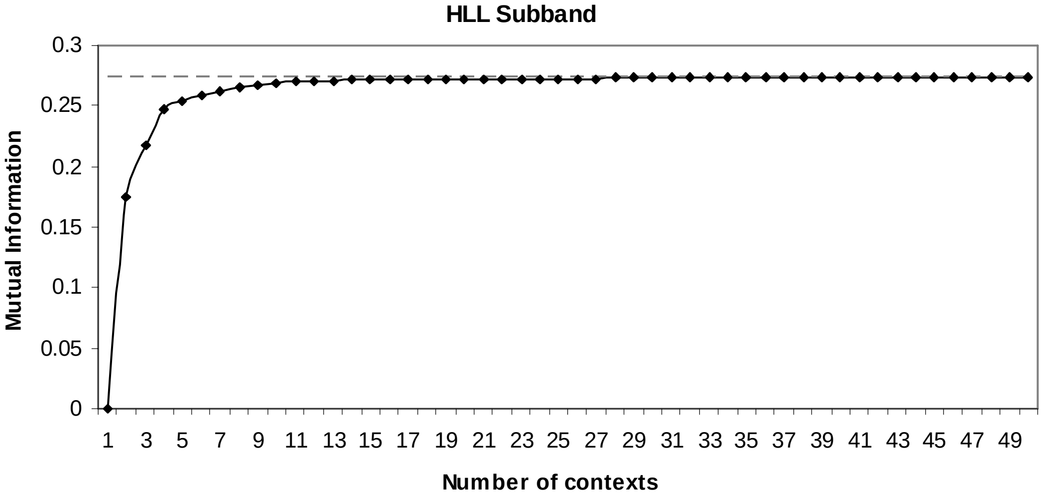

Figure 2. Optimal Mutual Information in relation to the number of context groups. The dashed line represents the maximal mutual information when having 2^{26} contexts.

The proposed algorithm starts with only one context group, and splits it at the first iteration. At subsequent iterations, the algorithm splits one of the available context groups into two, until the requested amount of groups is reached. The actual split point is determined by selecting the context classification candidate that maximizes the mutual information. As the total number of context groups F needed for a 3D context model is relatively close to 1 and far from I, this top-down method prevents joining the majority of contexts. Additionally, the number of searches needed to find the requested classification is now reduced to

$$F \cdot I \tag{2}$$

Using (1) and (2), it can be shown that for small F (i.e. $1 \leq F \leq 40$), and respectively large I, the bottom-up approach has a time complexity order of $O(I^2)$, while the proposed top-down approach in this paper has a complexity order of $O(I)$.

To reduce even further the amount of required split point searches, we introduce a relatively small search step size T (i.e. $T \leq 400$), to lower the amount of points that need to be tested. The split point candidate that is found is then re-evaluated together with the other candidates located in the surrounding range $[-T,T]$. This new re-evaluation step is performed with a reduced step size $T/2$. The process is then repeated until $T=1$, such that the final split point that maximizes the mutual information is found. The only disadvantage of this approach is that by using a step size T, it is possible that global maxima are missed, such that the search sometimes returns local maxima. Hence, we minimize the probability of having a local maximum contribution to the final context classification by splitting into more contexts groups, F_M, than those specified by F (i.e. $I >> F_M > F$). The context groups in the over-split context classification are then merged (top-down), again while maximizing the mutual information, until the total number of context groups is equal to F. By doing so, “bad” split points disappear from the classification, as their MI contribution is evidently less optimal than that of the other points. In theory, it is possible that a local maxima split point still finds it way into the final classification. In practice however, this does not cause any significant performance drawbacks, as the probability of this occurrence is very low and such local maxima often lie very close to the nearby global maxima.

With the procedure defined above, we are able to generate an optimized 3D context classification scheme that is well within reach of the computational limitations of today hardware.

3.2 Near-optimal context grouping

We have shown in the previous sub-section that the construction of a context model requires a Joint Probability Matrix.

Table 1. Obtained bits per pixel after lossless compression, using the 5x3 wavelet kernel.

5x3	bit depth	JPEG2000	2D	3D	New 3D-9	New 3D-16
MRINormalBrain	12	4.75	4.08	4.08	3.77	3.76
CTSpiral	12	5.82	5.38	5.44	5.07	5.06
CTNormalChest	12	5.17	5.02	5.07	4.72	4.72
AxialCT	12	4.09	3.84	3.89	3.65	3.65
UltrasoundSpine	8	5.44	4.84	4.95	4.53	4.61
3DPET	15	10.27	8.81	8.85	7.86	7.84
3DEcho	8	4.30	3.81	3.83	3.54	3.53

Table 2. Obtained bits per pixel after near-lossless compression, using the 9x7 wavelet kernel and performing no truncation.

9x7	bit depth	2D	3D	New 3D-9	New 3D-16
MRINormalBrain	12	3.79	3.79	3.45	3.44
CTSpiral	12	5.16	5.26	4.85	4.84
CTNormalChest	12	4.96	5.03	4.67	4.66
AxialCT	12	3.52	3.63	3.35	3.35
UltrasoundSpine	8	4.46	4.63	4.14	4.25
3DPET	15	8.49	8.52	7.53	7.50
3DEcho	8	3.54	3.56	3.23	3.23

The JPM used in our tests is based on a training data set that consists of nine volumetric images, wherein each image contributes to a series of probabilities which we have calculated for a set of bitrates.

The latter set includes a (near-)lossless bitrate and four lossy bitrates, respectively. Additionally, in the calculation of the JPM we have employed both the 5x3 and the 9x7 wavelet kernels. This approach allows us to take into consideration the statistical changes over different compression ratios and wavelet kernels, respectively. Furthermore, we should point out that the aim of this work is create a generic context model targeted at the compression of medical image data. In this respect, we have included CT, PET, MRI and ultrasound images in the training set.

The JPM computed with the compression settings described above makes it possible to calculate the optimally obtained mutual information in terms of the number of context groups. Thus, we illustrate in Figure 2, for the HLL subband, the relation between the optimal MI and the number of contexts. Mutual information values are computed as averages over the test set of data. Similar results are obtained for the other wavelet subbands.

It can be seen from Figure 2 that the increment of mutual information decreases as more context groups are used. It is clear that for 15 context groups the MI is almost equal to the maximal MI with 2^{26} contexts. Hence, for $F = 15$, the context model reaches near-optimality. For the purpose of this work, we have designed two new 3D context models; one contains 16 context groups (i.e. $F = 15$), while the other has only 9 (i.e. $F = 8$). Note that for JPEG2000 the context group with label 0 is a special context group that contains only the zero context (i.e. all significance states are zero), hence the values used for F. We have considered the latter context model (i.e. with $F = 8$) due to the fact that it uses the same number of context groups as the 2D JPEG2000 context model. In this respect, having the number of context groups for our 3D model equal to that of the 2D model allows for the simplification of codec implementations, supporting both 2D and 3D codestreams.

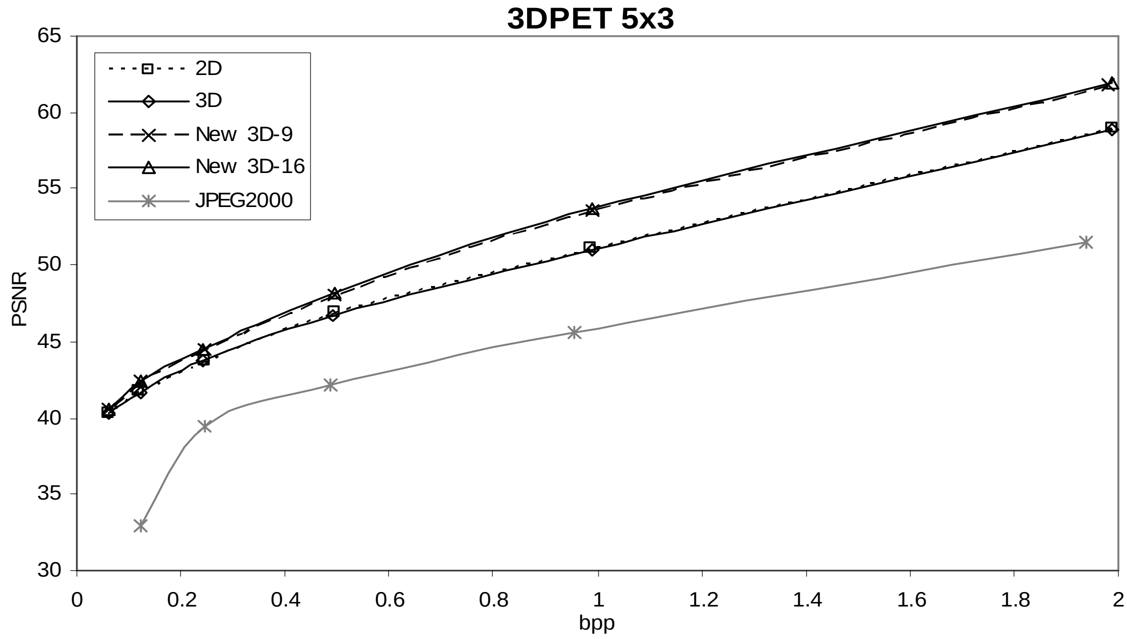

Figure 3. For 3DPET, using the 5x3 kernel, the PSNR values for different target bitrates, for the four different context models and for JPEG2000 slice-by-slice.

4 EXPERIMENTAL RESULTS

In order to allow an objective comparison between different context models, for each model investigated, we have compressed a set of volumetric datasets using the same codec and settings. The tests were performed using the 5x3 and 9x7 wavelet kernels, and targeted a predetermined set of bitrates and full lossless coding (i.e. not truncating the codestream), respectively. Five wavelet decompositions were performed in all three directions, except for the 'CTNormalChest' and '3DPET' datasets where, due to their small volumetric depth, only three decompositions were done in the Z direction. Please note that, for the 5x3 lossless coding, the PSNR is always infinity. However, since the 9x7 kernel is irreversible, the reported PSNR is in this case finite and thus the compression is at best near-lossless. The set of targeted bitrates include 2bpp, 1bpp, 0.5bpp, 0.25bpp, 0.125bpp and finally 0.0625bpp. The standard 2D JPEG2000 context model was used as the base reference and tested against the 3D context model of [5] and our two new 3D context models, respectively.

We illustrate in Table 1 and Table 2, for seven volumetric datasets, the lossless compression results obtained using the 5x3 and 9x7 wavelet kernels. We show the bitrates (in bpp) obtained using our JP3D codec with the proposed 3D context models with nine (3D-9) and sixteen (3D-16) context groups respectively. For the purpose of comparison, compression results for the JPEG2000 2D context model and the 3D context model of [5], when used with our JP3D codec, are also shown. Additionally, for the 5x3 wavelet kernel, we also show the results obtained using the Kakadu JPEG2000 Part 1 codec [9] with comparable compression parameters. The numbers obtained with the JPEG2000 codec are the result of a slice by slice compression. It can be seen that in all cases our proposed 3D-9 and 3D-16 context models outperforms by 5% to 10% both the existing 2D and 3D models for full-lossless coding. However, the difference in compression performance between the 3D-9 and the 3D-16 context model appears to be very moderate (less than 0.5%). This is expected as it matches the observation made in Figure 2. The optimally obtained MI, with both 8 and 15 context groups respectively, approaches the maximal MI closely.

Furthermore, we plot in Figure 3, Figure 4 and Figure 5 the compression results obtained on the '3DPET', 'AxialCT' and the ' MRINormalBrain' datasets, at different target bitrates and for different codecs and context models. For these experiments, we have employed the 5x3 wavelet kernel with five decompositions in the X and Y directions respectively. Additionally, for the tests with the JP3D codec, also five decompositions in the Z direction were performed on the 'AxialCT' and the ' MRINormalBrain' data sets and three on the '3DPET' data set.

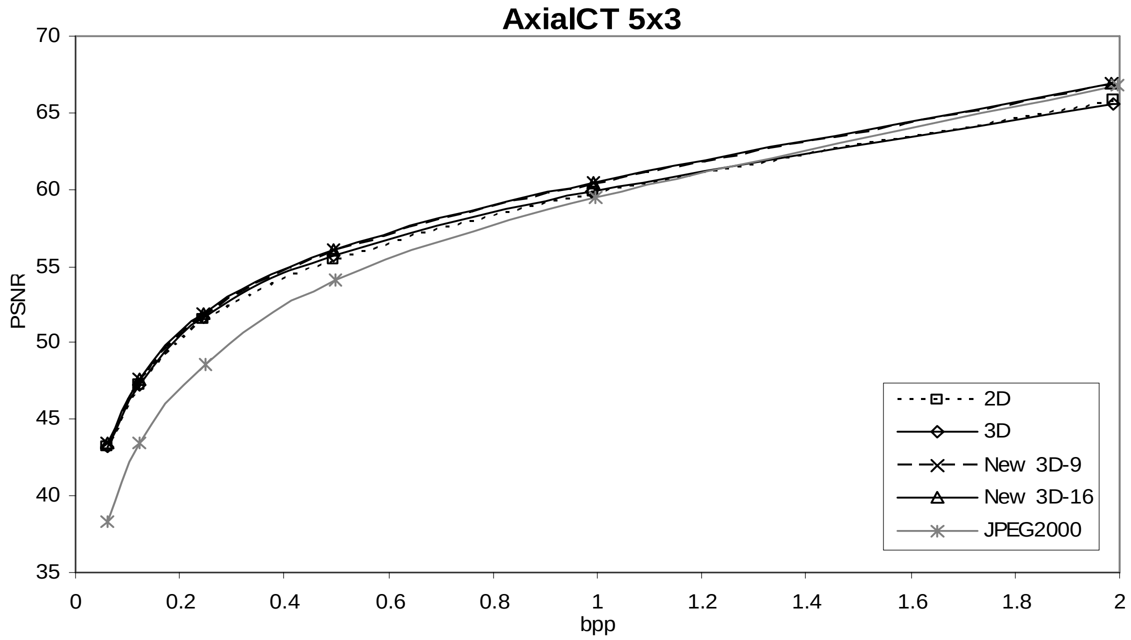

Figure 4. For AxialCT, using the 5x3 kernel, the PSNR values for different target bitrates, for the four different context models and for JPEG2000 slice-by-slice.

For each codec, we plot the real bitrate of the compressed codestream and the corresponding Peak Signal to Noise Ratio (PSNR) value. We can conclude from these figures that for all the investigated bitrates, the new proposed context model clearly performs better than the current standard 2D context model and the context model of [4], respectively. In this sense, we report PSNR gains of up to 3dB. Moreover, it can be seen that the compression performance gain increases as one goes to higher bitrates. This is of paramount importance in medical imaging applications, for which high quality is a primary concern. At very low bitrates the difference in compression performance is less obvious or almost non existent when comparing the four JP3D context models. This is due to the fact that when collecting the statistical data of the training data sets, used to construct the joint probability matrix, a tradeoff between optimization for low or high bitrate targets had to be made. Because, it is often important that compressed medical data sets can be reconstructed with high or even lossless image quality, we choose to benefit the higher bitrate targets. Logically, this resulted into a context model that performs better at the high bitrate targets.

5 CONCLUSIONS

We have proposed in this paper a context design algorithm that allows a fast design of context models for the EBCOT coder. The algorithm comes as a solution for the situations in which the number of allowable initial contexts is very large, while keeping complexity and execution time well within the limits of today's hardware. In this sense, the three-dimensional 3x3x3 context neighborhood investigated in this paper is already an example of an instantiation that would have otherwise been computationally unfeasible.

Furthermore, we have designed a new 3D context model for the 3D extension of JPEG2000 (JPEG2000 Part 10 – JP3D). We show that the model improves the lossless compression efficiency by 5% to 10% on all the tested 3D datasets. Finally, in lossy coding, for the same test set, for a wide range of target bitrates, we report PSNR gains of up to 3dB.

6 FUTURE WORK

This paper provides a working 3D context model for JPEG2000 that compression-wise significantly outperforms the standardized 2D context model [1] or any other context model thus far tested. Future research will focus on representing the derived context model in a usable tabular format, i.e. as defined in JPEG2000 Part 1 for the two-dimensional case.

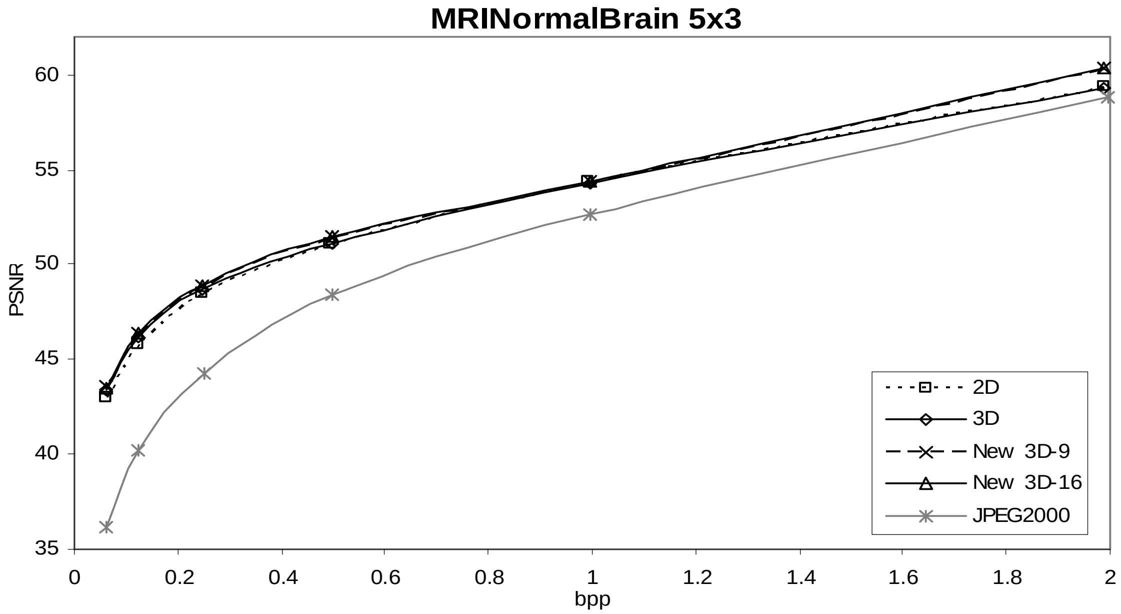

Figure 5. For MRINormalBrain, using the 5x3 kernel, the PSNR values for different target bitrates, for the four different context models and for JPEG2000 slice-by-slice.

This representation will allow an easy classification of the resulting contexts into context labels employed by the arithmetic encoder. We envisage that this approach will possibly result in a slightly less optimal compression performance than when using the full classification map, but will nonetheless remain competitive with respect to the latter.

7 ACKNOWLEDGEMENTS

This work was supported by the Interdisciplinair Instituut voor Breedbandtechnologie (IBBT) (E-HIP Project) and the Fund for Scientific Research Flanders (FWO) (via the postdoctoral fellowship of Adrian Munteanu and Peter Schelkens).

REFERENCES

1. "Information Technology - JPEG 2000 Image Coding System: Part 1 - Core Coding System," ISO/IEC JTC1/SC29/WG1 IS 15444-1, 2000.
2. "Information Technology - JPEG 2000 Image Coding System: Part 2 - Extensions," ISO/IEC JTC1/SC29/WG1 IS 15444-2, 2002.
3. "Information Technology - JPEG 2000 Image Coding System: Part 10 - Extensions for Three-Dimensional Data (JP3D) - FCD V1.0," ISO/IEC JTC1/SC29/WG1 N4101, 2006.
4. P. Schelkens, A. Munteanu, J. Barbarien, M. Galca, X. Giro i Nieto, and J. Cornelis, "Wavelet Coding of Volumetric Medical Datasets," *IEEE Transactions on Medical Imaging*, vol. 22, pp. 441-458, 2003.
5. D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Transactions on Image Processing*, vol. 9, pp. 1158-1170, 2000.
6. J. Liu and P. Moulin, "Information-Theoretic Analysis of Interscale and Intrascale Dependencies between Image Wavelet Coefficients," *IEEE Transactions on Image processing*, vol. 10, pp. 1647-1658, 2001.
7. Z. Liu and L. J. Karam, "Mutual Information-Based Analysis of JPEG2000 Contexts," *IEEE Transactions*, vol. 14, pp. 411-422, 2005.
8. T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NT, USA: Wiley, 1991.
9. D. Taubman, "Kakadu Software," http://www.kakadusoftware.com/.