A rate-distortion study on microarray image compression

Luís M. O. Matos
luismatos@ua.pt
António J. R. Neves
an@ua.pt
Armando J. Pinho
ap@ua.pt

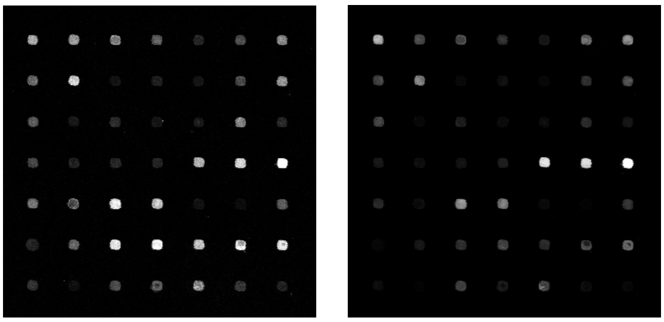
Signal Processing Laboratory
IEETA/DETI
University of Aveiro,
3810-193 Aveiro, Portugal

Abstract

In this paper we present a rate-distortion study on microarray image compression. We evaluate two image coding standards (JBIG and JPEG2000), a method based on bitplane decomposition, and a method recently proposed based on binary-tree decomposition. According to the results obtained JPEG2000 attains the worse compression results particularly for higher bitrates. On the other hand, the method based on binary-tree decomposition outperformed all the other compression methods in terms of L2-norm and L_∞-norm. The results attained for the binary-tree decomposition method can be explained due to its nature, consisting on coding each node of the tree with the goal of minimizing the L_∞ error.

1 Introduction

The raw data resulting from a microarray experiment consist of a pair of 16 bits per pixels of grayscale images (see Figure 1). These images, depend on the size of the array and the resolution of the scanner, and may require tens of megabytes to be stored or transmitted without any compression loss. Due to this fact, and the need for long-term storage and efficient transmission, the development of lossless compression methods with progressive decoding capabilities is an important challenge. Moreover, this progressive decoding capability is useful because the decoding process can be stopped at any time obtaining an intermediate image with some loss. This loss control can be based on the same principles that are used to extract genetic information from the microarrays.



Channel 1 Channel 2

Figure 1: An example of a microarray experiment (crop portion of "Def661" image from the ISREC dataset).

2 Related work and goals

In the literature we can find several lossless compression methods for microarrays images [1, 3, 4]. One interesting characteristic of some of these methods is its capability of progressive decoding the compressed image. This characteristic is very important because we can stop the decoding process at any time obtaining a partial image with some loss, loss that will not degrade the genetic information that will be extracted later. Besides that, the progressive decoding approach allows us to also decode the original image without any loss. This is also very important because the analysis techniques are subject to changes which means that in the future we may require to have the original image, without discarding any information.

Recently, we have proposed a lossless compression method for microarray images, based on a hierarchical organization of the intensity levels followed by finite-context modeling [2]. The organization of the intensity levels is attained by means of a binary tree. This binary tree is build and traversed in a way that each node that is expanded, is selected taking into account the smallest L_∞ reconstruction error.

Neves and Pinho proposed in 2009 a compression method based on a bitplane decomposition approach that uses a 3D finite-context model to drive the arithmetic coder [4]. Their work was inspired by EIDAC [5] but their method was improved by using image-dependent context models that are build with the goal of finding the "best" context configuration to encode each bitplane.

The goal of this research work is to evaluate in terms of rate-distortion this method based on a binary-tree decomposition [2], the method presented in 2009 by Neves and Pinho [4], and also two image coding standards (JBIG and JPEG2000) that have support for lossy-to-lossless compression.

3 Experimental results

The goal of this work is to evaluate the performance of four different methods in terms of rate-distortion. Two of the methods are specific for microarray compression [2, 4]. The other two methods are the JBIG and JPEG2000 standards.

In the literature, we can find several rate-distortion metrics that can be used to measure the performance of compression algorithms. In this research work we decided to use the root mean square error (or L2) and the maximum absolute error (L_∞). Because of the lack of space, we chose only two images to evaluate the performance of the methods mention earlier. One image of the Apo AI data set ("1230cG") and one image of the Microzip data set ("array1"). For the other data sets and images, the obtained results are quite similar so the conclusions made for these two images are the same as for the remaining images.

The rate-distortion curves of the two selected images can be found in Figs 2 and 3. The left charts are related to the L2-norm (root mean squared error) and the right ones were plotted in terms of L_∞-norm (maximum absolute error). Regarding the first metric (L2-norm), we can notice that lossless JPEG2000 provides similar rate-distortion results for lower bitrates (lower than 8 bpp). The poor results in higher bitrates for JPEG2000 could be explained by the default parameters used in JPEG2000. These default parameter values could not be suited for this kind of images with 16 bits per pixel and their particular characteristics in terms of noise, histogram, etc. Regarding the L_∞-norm, we observed that JPEG2000 has in fact the worst rate-distortion results. Similar to the previous metric, we can see that the rate-distortion results for JPEG2000 suffer a sudden deviation for higher bitrates which is probably related to the same problem pointed out earlier. When comparing methods [4] and [2], we can see that method [2] provides better rate-distortion results when compared to method [4]. The main reason to the previous statement is due to the nature of method [2]. As mentioned earlier in Section 2, method [2] has in its core a mechanism that minimizes the L_∞ error along the encoding/decoding process. This allows the method to obtain better rate-distortion results. Furthermore, method [4] was not designed with an error minimization goal in mind. It processes each bitplane of the microarray image without looking to any kind of error metric. The error after processing each bitplane depends on the remaning information that was not yet processed (on the lower bitplanes).

4 Conclusions

The goal of this study was to compare, in terms of rate-distortion, several compression methods. We used two image coding standards (JBIG and JPEG2000) and two different specialized methods for microarray images [2, 4]. According to the obtained rate-distortion results, we con-

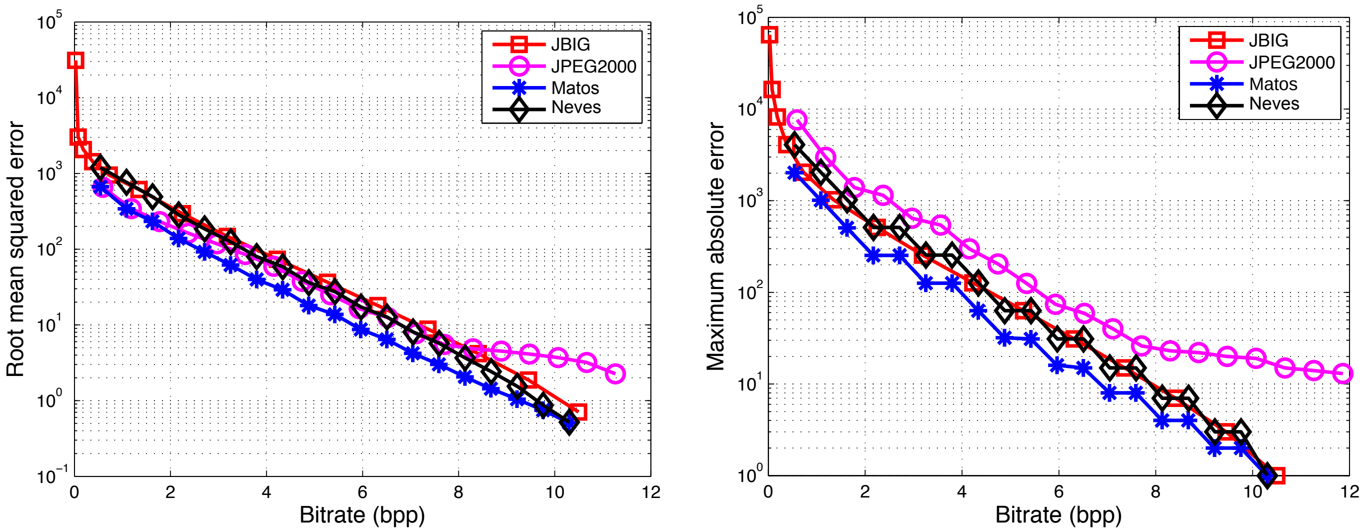


Figure 2: Rate-distortion curves for methods [2, 4], JBIG, and JPEG200, regarding image "1230c1G" of the Apo AI data set. Results are given in terms of L2-norm (root mean squared error) on the left chart and in terms of L_∞-norm (maximum absolute error) on the right chart. The curves indicated as "Matos" correspond to method [2], whereas the curves indicated as "Neves" correspond to method [4].

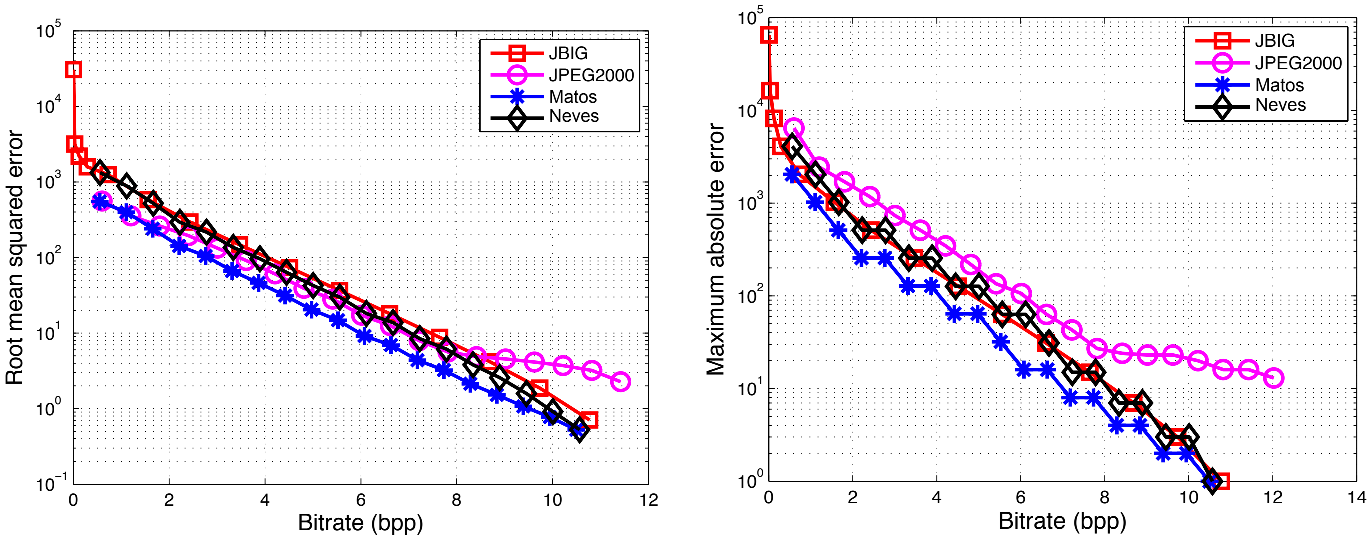


Figure 3: Rate-distortion curves for methods [2, 4], JBIG, and JPEG200, regarding image "array1" of the Microzip data set. Results are given in terms of L2-norm (root mean squared error) on the left chart and in terms of L_∞-norm (maximum absolute error) on the right chart. The curves indicated as "Matos" correspond to method [2], whereas the curves indicated as "Neves" correspond to method [4].

cluded that among the methods used in this work, JPEG2000 attained similar results to the other methods for lower bitrates (lower than 8 bpp), in terms of root mean squared error. The poor performance for higher bitrates is probably due to the default parameters used in the JPEG2000 codec. Regarding the other measure, the maximum absolute error, the JPEG2000 standard continues to attain the worse results when compared to the other method, mainly for higher bitrates. The phenomenon is probably caused by the same issue mention earlier. Regarding the specialized methods, the obtained results in both rate-distortion metrics are better when compared to JBIG and JPEG2000. In fact, the method based on binary-tree decomposition outperforms the other three methods due to its nature. The core method process each node of the tree in a way to minimize the L_∞ error.

Funding

This work was partially supported by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Foundation for Science and Technology, in the context of a PhD Grant (FCT reference SFRH/BD/86531/2012) and a project (FCT reference PEst-OE/EEI/UI0127/2014).

References

[1] S. Battiato and F. Rundo. A bio-inspired CNN with re-indexing engine for lossless DNA microarray compression and segmentation. In *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2009*, volume 1-6, pages 1737–1740, Cairo, Egypt, November 2009. doi: 10.1109/ICIP.2009.5413629.

[2] L. M. O. Matos, A. J. R. Neves, and A. J. Pinho. Compression of microarray images using a binary tree decomposition. In *Proceedings of the 22nd European Signal Processing Conference, EUSIPCO 2014*, September 2014.

[3] A. Neekabadi, S. Samavi, S. A. Razavi, N. Karimi, and S. Shirani. Lossless microarray image compression using region based predictors. In *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2007*, volume 2, pages 349–352, San Antonio, Texas, USA, September 2007.

[4] A. J. R. Neves and A. J. Pinho. Lossless compression of microarray images using image-dependent finite-context models. *IEEE Transactions on Medical Imaging*, 28(2):194–201, February 2009.

[5] Y. Yoo, Y. G. Kwon, and A. Ortega. Embedded image-domain compression using context models. In *Proceedings of the EEE Int. Conf. on Image Processing, ICIP-99*, volume I, pages 477–481, Kobe, Japan, October 1999.