# A BITPLANE BASED ALGORITHM FOR LOSSLESS COMPRESSION OF MICROARRAY IMAGES


*António J. R. Neves and Armando J. Pinho*

Dept. de Electrónica, Telecomunicações e Informática / IEETA
Universidade de Aveiro, 3810–193 Aveiro, Portugal
an@ua.pt, ap@ua.pt



## ABSTRACT

The use of microarray expression data in state-of-the-art biology has been well established. The widespread adoption of this technology coupled with the significant volume of data generated per experiment, in the form of images, have led to significant challenges in storage and query-retrieval of the data from microarray experiments. In this paper, we present a lossless bit-plane based method for efficient compression of microarray images. This method is based on arithmetic coding using a 3D context model. We also present an extension to the proposed method to show the limit of using 3D context models for encoding this type of images. Our methods produce an embedded bit-stream that allows progressive, lossy-to-lossless decoding. We compare the compression efficiency of the proposed method with JBIG and with the two most recent specialized methods. The proposed method gives better results for all images of the test sets and confirms the effectiveness of bit-plane based methods and finite-context models for lossless compression of microarray images.


## 1. INTRODUCTION

The DNA microarray technology has become an important tool in the study of gene function, regulation, and interaction across large numbers of genes, and even entire genomes. It allows the analysis of thousands of genes in a single experience [1, 2].

The raw data of a microarray experiment consist of a pair of 16 bits per pixel grayscale images. These images are analyzed using a variety of software tools which extract relevant information, such as the intensity of the spots and the background level. This information is then used to evaluate the expression level of individual genes [1, 2]. Depending on the size of the array and the resolution of the scanner, these images may require several tens of megabytes in order to be stored or transmitted.

Microarrays have been the focus of significant research. Most of this effort has been directed to the analysis of the data resulting from such experiments, whereas problems such as the efficient representation of the microarray images has received relatively less attention. However, giving the massive amount of data currently produced and the need of long-term storage and efficient transmission, the development of efficient compression methods is an important challenge.

The common approach towards the compression of microarray images has been based on image analysis for spot finding (griding followed by segmentation) with the aim of separating the microarray image data into channels based on pixel similarities [3, 4, 5, 6, 7, 8, 9, 10, 11]. Once separated, the channels are compressed individually, together with the segmentation information. Although appealing, image dependent compression methods may potentially run into trouble if the assumptions in which they are based change in the future. One such assumption might be, for example, the rectangular organization of the spots. In fact, although initially this was the organization used for spot placement in the microarrays, more recently, other non-rectangular packings are being used.

To our knowledge, the technique that we propose in this paper is the best one currently available in terms of compression efficiency of microarray images. The method is based on arithmetic coding driven by a 3D finite-context model (preliminary results have been presented in [12]). Basically, the image is compressed on a bit-plane basis, going from the most significant bit-plane to the least significant bit-plane. The finite-context model used by the arithmetic encoder uses (causal) pixels from the bit-plane under compression and also pixels from the bit-planes already encoded. In this paper, we extended this method in order to find the best context shape for each bit-plane and we provide detailed results showing the effectiveness of our methods. Moreover, we provide detailed results of our method comparing them with JBIG [13, 14] and also with two recent specialized methods, MicroZip [10] and Zhang's method [11].

## 2. THE PROPOSED METHOD

In this paper, we present a lossless compression method for microarray images based on an adaptive finite-context model followed by arithmetic coding. This method was inspired by EIDAC [15], a compression method that has been used with success for coding images with a reduced number of intensities (simple images). Here, we show that similar ideas can also be used for developing an efficient compressing method for microarray images. The original method proposed by Yoo *et al.* could not be used in these images due to its properties (they have 16 planes and a large number of different intensities).

In the proposed method, the images are compressed on a bit-plane basis, starting from the most significant bit-plane and stopping at the least significant bit-plane or whenever a bit-plane requires more than one bit per pixel for encoding. In this case, the rest of the bit-planes are left uncoded. The causal finite-context model that drives the arithmetic encoder uses pixels both from the bit-plane currently being encoded and from the bit-planes already encoded (see Fig. 1).

When encoding the eight least significant bit-planes, the finite-context model is only formed with pixels from the upper bit-planes. This procedure avoids the degradation in compression rate which occurs because, in general, the eight least significant bit-planes are close to random and, therefore, they are almost incompressible [6]. Moreover, we have introduced another mechanism that helps overcoming this problem. As the method proceeds encoding the image, the average bit-rate obtained after encoding each bit-plane is mon-

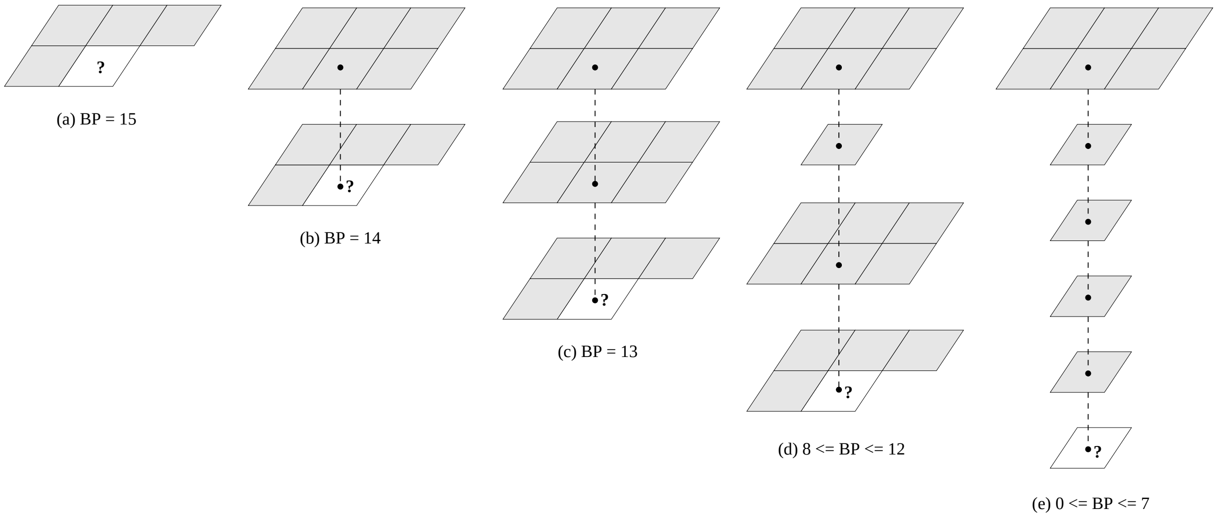


**Fig. 1**. Context configurations used by the proposed method at five different compression stages: (a) when encoding the most significant bit-plane (four pixels of context); (b) when encoding the second most significant bit-plane (ten pixels of context); (c) when encoding the third most significant bit-plane (16 pixels of context); (d) from the fourth until the eighth most significant bit-planes (17–21 pixels of context); (e) the eight least significant bit-planes (13–20 pixels of context).

itored. If, for some bit-plane, the average bit-rate exceeds one bit per pixel, then we stop the encoding process and the remaining bit-planes are saved without compression.

## 3. EXPERIMENTAL RESULTS

The compression method proposed in this paper was evaluated using microarray images that have been collected from three different publicly available sources: (1) 32 images that we refer to as the Apo AI set and which have been collected from `http://www.stat.berkeley.edu/users/terry/zarray/Html/index.html` (this set was previously used by Jörnsten *et al.* [4, 6]); (2) 14 images forming the ISREC set which have been collected from `http://www.isrec.isb-sib.ch/DEA/module8/P5_chip_image/images/`; (3) three images previously used to test MicroZip[10] and Zhang's method[11], which were collected from `http://www.cs.ucr.edu/~yuluo/MicroZip/`. These three sets have also been used in the experiments reported in [16].

Image size ranges from $1000 \times 1000$ to $5496 \times 1956$ pixels, i.e., from uncompressed sizes of about 2 MBytes to more than 20 MBytes (all images have 16 bits per pixel). The average results presented take into account the different sizes of the images, i.e., they correspond to the total number of bits divided by the total number of image pixels.

Table 1 shows the average compression results, in bits per pixel, for the three sets of images described above. In this table, we also show the effect of using Gray codes instead of the natural binary codes for representing the pixel intensities. Gray coding has been proposed as a means of improving the compression attained by JBIG in graylevel images [17]. Indeed, as can be seen in Table 1, the compression results provided by JBIG improved, on average, about $3.6\%$ when using Gray codes. However, for the ISREC image set, it resulted in a small loss of performance. On the other hand, the proposed method was quite insensible to the use of Gray codes. On average, it improved less than $0.2\%$, which can be justified by its exploitation of inter-bit-plane dependencies. Unlike JBIG, this (small) improvement was consistent: none of the image sets suffered a loss of compression performance when using Gray codes.

Compared to JBIG, the proposed method provided an overall compression gain of about $5.6\%$. Table 2 confirms the performance of our method relatively to two recent specialized methods for compressing microarray images: MicroZip [10] and Zhang's method [11]. As can be observed, the method that we propose provides compression gains of $5.1\%$ relatively to JBIG, $7.4\%$ relatively to MicroZip and $4.5\%$ in relation to Zhang's method, on a set of test images that has been used by all methods.

**Table 1**. Average compression results, in bits per pixel, using JBIG and the proposed method, combined with Gray codes and the natural binary codes.

| Image Set | JBIG | | Proposed | |
|---|---|---|---|---|
| | Bin Code | Gray Code | Bin Code | Gray Code |
| Apo_AI | 11.367 | 10.851 | 10.280 | 10.258 |
| ISREC | 10.849 | 10.925 | 10.199 | 10.194 |
| MicroZip | 9.788 | 9.297 | 8.840 | 8.826 |
| **Average** | **10.783** | **10.393** | **9.826** | **9.809** |

**Table 2**. Compression results, in bits per pixel, using JBIG, MicroZip, Zhang's method and the proposed method.

| Image | JBIG | MicroZip | Zhang | Proposed |
|---|---|---|---|---|
| array1 | 11.819 | 11.490 | 11.380 | 11.105 |
| array2 | 9.071 | 9.570 | 9.260 | 8.628 |
| array3 | 8.351 | 8.470 | 8.120 | 7.962 |
| **Average** | **9.297** | **9.532** | **9.243** | **8.826** |

## 4. CONCLUSION

In this paper, we presented an efficient method for lossless compression of microarray images, allowing progressive, lossy-to-lossless decoding. This method is based on bit-plane compression using finite-context models and arithmetic coding. It does not require griding and / or segmentation as most of the specialized methods that have been proposed do. This may be an advantage if only compression is sought, since it reduces the computational complexity of the method. Moreover, since it does not depend on image content, it is robust, for example, against layout changes in spot placement.

The results obtained have been compared with JBIG [13] and with two recent specialized methods: MicroZip [10] and Zhang's method [11]. The results show that the proposed method has better compression performance than JBIG in all test sets. This is not surprising, since, contrarily to JBIG, it explores inter-bit-plane dependencies. Compared to the specialized methods, the proposed method was the best for all images in the MicroZip test set, showing lossless compression gains of more than $4.5\%$.

## 5. REFERENCES


[1] S. K. Moore, "Making chips to probe genes," *IEEE Spectrum*, vol. 38, no. 3, pp. 54–60, Mar. 2001.

[2] P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. Earle-Hughes, E. Snesrud, N. Lee, and J. Q., "A concise guide to cDNA microarray analysis," *Biotechniques*, vol. 29, no. 3, pp. 548–562, Sept. 2000.

[3] R. Jörnsten, B. Yu, W. Wang, and K. Ramchandran, "Compression of cDNA and inkjet microarray images," in *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2002*, vol. 3, Rochester, NY, Sept. 2002, pp. 961–964.

[4] R. Jörnsten, Y. Vardi, and C.-H. Zhang, "On the bitplane compression of microarray images," in *Proc. of the 4th Int. L1-norm Conf.*, Y. Dodge, Ed., 2002.

[5] J. Hua, Z. Xiong, Q. Wu, and K. Castleman, "Fast segmentation and lossy-to-lossless compression of DNA microarray images," in *Proc. of the Workshop on Genomic Signal Processing and Statistics, GENSIPS*, Raleigh, NC, Oct. 2002.

[6] R. Jörnsten, W. Wang, B. Yu, and K. Ramchandran, "Microarray image compression: SLOCO and the effect of information loss," *Signal Processing*, vol. 83, pp. 859–869, 2003.

[7] J. Hua, Z. Liu, Z. Xiong, Q. Wu, and K. Castleman, "Microarray BASICA: background adjustment, segmentation, image compression and analysis of microarray images," in *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2003*, vol. 1, Barcelona, Spain, Sept. 2003, pp. 585–588.

[8] N. Faramarzpour, S. Shirani, and J. Bondy, "Lossless DNA microarray image compression," in *Proc. of the 37th Asilomar Conf. on Signals, Systems, and Computers, 2003*, vol. 2, Nov. 2003, pp. 1501–1504.

[9] N. Faramarzpour and S. Shirani, "Lossless and lossy compression of DNA microarray images," in *Proc. of the Data Compression Conf., DCC-2004*, Snowbird, Utah, Mar. 2004, p. 538.

[10] S. Lonardi and Y. Luo, "Gridding and compression of microarray images," in *Proc. of the IEEE Computational Systems Bioinformatics Conference, CSB-2004*, Stanford, CA, Aug. 2004.

[11] Y. Zhang, R. Parthe, and D. Adjeroh, "Lossless compression of DNA microarray images," in *Proc. of the IEEE Computational Systems Bioinformatics Conference, CSB-2005*, Stanford, CA, Aug. 2005.

[12] A. J. R. Neves and A. J. Pinho, "Lossless compression of microarray images," in *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-2006*, Atlanta, GA, Oct. 2006, pp. 2505–2508.

[13] ISO/IEC, *Information technology - Coded representation of picture and audio information - progressive bi-level image compression*, International Standard ISO/IEC 11544 and ITU-T Recommendation T.82, Mar. 1993.

[14] D. Salomon, *Data compression - The complete reference*, 2nd ed. Springer, 2000.

[15] Y. Yoo, Y. G. Kwon, and A. Ortega, "Embedded image-domain compression using context models," in *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-99*, vol. I, Kobe, Japan, Oct. 1999, pp. 477–481.

[16] A. J. Pinho, A. R. C. Paiva, and A. J. R. Neves, "On the use of standards for microarray lossless image compression," *IEEE Trans. on Biomedical Engineering*, vol. 53, no. 3, pp. 563–566, Mar. 2006.

[17] M. Abdat and M. G. Bellanger, "Combining Gray coding and JBIG for lossless image compression," in *Proc. of the IEEE Int. Conf. on Image Processing, ICIP-94*, vol. III, Austin, TX, Nov. 1994, pp. 851–855.