Lossless compression of symbolic and biomedical images using finite-context models

Luis M. O. Matos, António J. R. Neves, Diogo Pratas, and Armando J. Pinho
Department of Electronics, Telecommunication and Informatics
Institute of Electronics and Telematics Engineering of Aveiro
University of Aveiro

ABSTRACT
During the last years, for most medical imaging modalities, the amount of data produced is increasing exponentially. The same is also true for DNA sequencing, where impressive progress has been made in the last decade. This progress caused a significant growth in the amount of data related to medical imaging and DNA sequencing that need to be stored and transmitted. In order to reduce the storage cost and transmission time, without jeopardizing the image quality, efficient and robust compression methods are required. In our research work, we developed compression methods based on finite-context models to compress microarray images and whole genome alignments, which is a particular voluminous dataset in genomics.

INTRODUCTION
The goal of our work is to develop sophisticated compression methods that can be used for compressing medical images and some DNA sequencing datasets. The core of our compression methods consists of an adaptive finite-context model followed by arithmetic coding (see Fig. 1).

RESULTS
Regarding the compression of microarray images, we developed two methods based on:
- Bitplane decomposition.
- Binary tree decomposition.
We performed some simulations in six datasets and the results obtained are depicted in Fig. 4.

CONCLUSION
In this poster, we present compression results for lossless compression of microarray images and MSA blocks (symbolic images). The core of the proposed methods consists of an adaptive finite-context model followed by arithmetic coding. For microarray images, we used two decomposition approaches: bitplane and binary tree decomposition. We compared the results with JBIG, JPEG-LS, and JPEG2000. On average, our method requires between 5.3 to 10.2 bits per pixel, depending on the dataset.

REFERENCES

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 a context of a PIDD Grant (FCT reference SFRH/BD/86531/2012) and the project FCOMP-01-0124-FEDER-020682 (FCT reference PEst-OE/EEI/UI0127/2011).