

Astronomical Image Compression Techniques Based on ACC and KLT Coder

J. Schindler, P. Páta, M. Klíma, K. Fliegel

Abstract

This paper deals with a compression of image data in applications in astronomy. Astronomical images have typical specific properties — high grayscale bit depth, size, noise occurrence and special processing algorithms. They belong to the class of scientific images. Their processing and compression is quite different from the classical approach of multimedia image processing. The database of images from BOOTES (Burst Observer and Optical Transient Exploring System) has been chosen as a source of the testing signal. BOOTES is a Czech-Spanish robotic telescope for observing AGN (active galactic nuclei) and the optical transient of GRB (gamma ray bursts) searching. This paper discusses an approach based on an analysis of statistical properties of image data. A comparison of two irrelevancy reduction methods is presented from a scientific (astrometric and photometric) point of view. The first method is based on a statistical approach, using the Karhunen-Loève transform (KLT) with uniform quantization in the spectral domain. The second technique is derived from wavelet decomposition with adaptive selection of used prediction coefficients. Finally, the comparison of three redundancy reduction methods is discussed. Multimedia format JPEG2000 and HCOMPRESS, designed especially for astronomical images, are compared with the new Astronomical Context Coder (ACC) coder based on adaptive median regression.

Keywords: astronomical image compression, Karhunen-Loève transform, dark frame compression, loss less compression algorithm, Astronomical Context Coder (ACC), JPEG 2000.

1 Introduction

This paper deals with scientific image data compression. The data for analysis was collected during work on the international (Czech-Spanish-Italian) BOOTES experiment (Burst Observer Optical Transient Exploring System) [2]. BOOTES has been in service since 1998 as the first Spanish robotic telescope for sky observation [4]. This system is one of three similar systems in full operation in the world, and has three main stations. The first one is located in the southern Spain (in Mazagon, near Huelva), and has been in full operation since July 1998. The first version of the system was completed in July 2001. The main aim of the project is to observe extragalactic objects and to detect a new optical transient (OT) of gamma ray burst (GRB) sources. BOOTES has been operated in very close co-operation with a satellite observation of the gamma and roentgen universe INTEGRAL satellite. INTEGRAL is an orbital astrophysics laboratory of the European Space Agency (ESA) and it has been in space since November 2002.

Due to the limited capacity of storage media, an efficient data compression algorithm has to be applied. Lossless compression algorithms are often used in scientific applications, but their efficiency is limited. The maximum achieved compression ratio depends above all on the data type and on the amount

of image signal entropy. The usual dictionary or entropy lossless algorithms are Run Length Encoding (RLE), Lempel Ziv Welch (LZW), Huffman or arithmetic coding. The typical compression ratios of these lossless algorithms are from 1 : 1.1 to 1 : 5 for astronomical images [1].

The second approach involves the use of compression techniques characterized by decorrelated parameters. Typical examples of this option are JPEG and JPEG2000 standards, but data impairment has to be taken into account in the case of lossy coding. It is necessary to consider whether algorithms optimized for multimedia applications and human vision are suitable for compressing scientific image data.

Astronomical image data stored in archives is often accessed later to perform a new study, new comparisons and measurements. It is not possible to fix a set of investigation methods which may be applied to the astronomical image in the future. It is therefore not possible to determine in advance an admissible loss of image information during the compression process. The best way to guarantee maximally accurate and reliable results from post-processing an astronomical image is to preserve the image without any change or loss of information. For this reason lossless compression techniques are often preferred in this area.

Recent lossy and lossless still image compression formats are powerful tools for compressing all kinds of

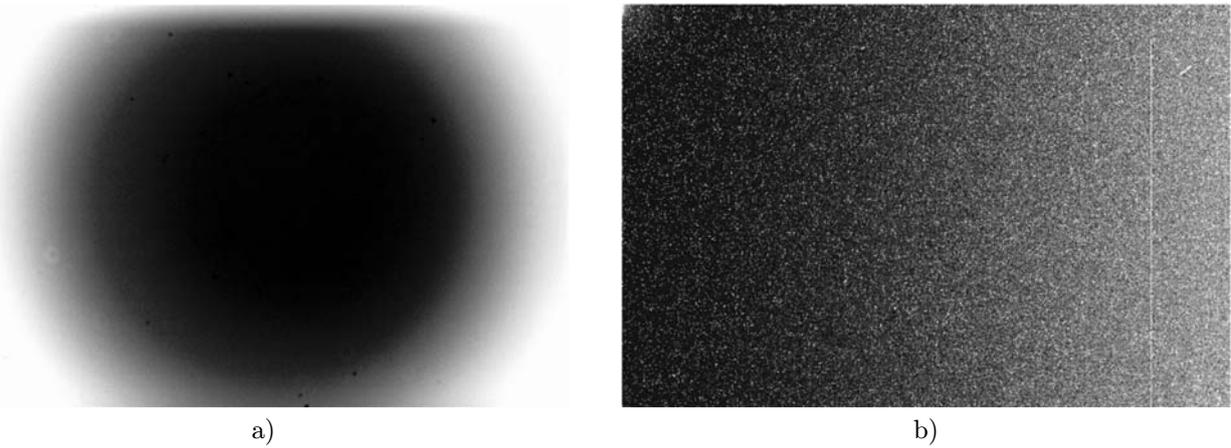


Fig. 1: Correction image data from the BOOTES project ($1024 \times 1536 \times 16$ bits) a) image for correcting non-uniform sensitivity of the whole detection system flat field (FF), b) map of the dark current of the CCD sensor — dark image (DI)

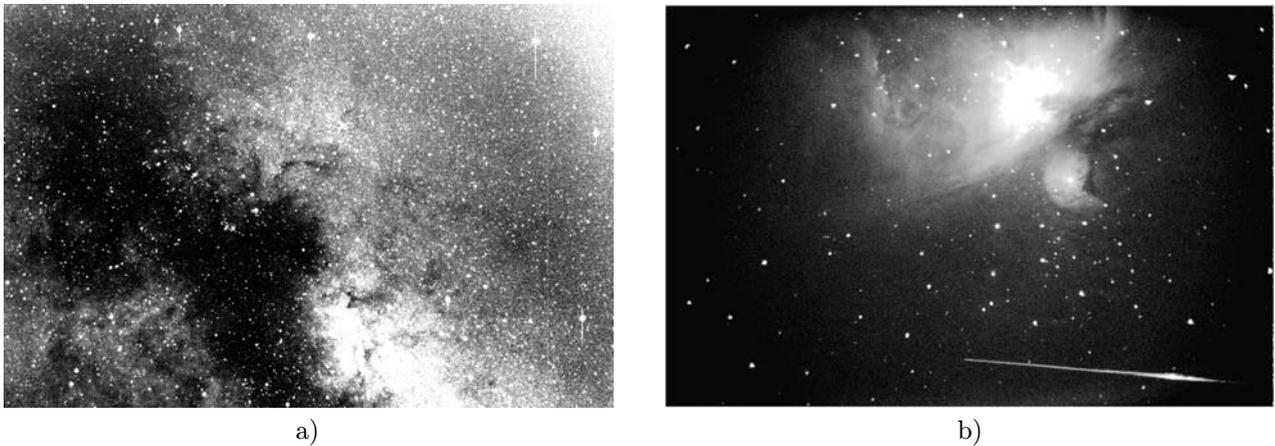


Fig. 2: Input image data from the BOOTES project ($1024 \times 1536 \times 16$ bits) a) Image from wide field camera M7 and Milky Way with many objects (size smaller than 10 pixels), b) M42 nebulae with a satellite tray. Image obtained from the DEEP SKY camera

common images (pictures, text, schemes, etc.). The performance of a compression algorithm generally depends on its ability to anticipate the image function of the processed image. In other words, a compression algorithm, in order to be successful, has to take fullest advantage of coded image properties. Astronomical data forms a special class of images that have general image properties, and also some specific characteristics. If a new coder is able to make correct use of knowledge of these special properties, this will lead to superior performance on this specific class of images, at least in terms of the compression ratio. Applying special compression algorithms based on specific properties of wavelet, fractal or Karhunen-Loève transform [9] seems to be a better solution for astronomical image data compression.

2 Astronomical images

The data coder has been optimized for four image types:

- image for correcting the non-uniform sensitivity of the whole detection system flat field (FF) (see Figure 1a). Note the shadow of the dust particle in the left part of the center of the image.
- a map of the dark current of the CCD sensor – dark frame (DF) (see Figure 1b). Note the bad CCD column in the right part of image.
- light images (LI) from wide and ultra-wide field cameras (EQ focus length shorter than 100 mm) (see Figure 2a). The size of the objects (especially stars) does not exceed 10 square pixels.
- light image with high spatial resolution — deep sky images (DSL) (see Figure 2b).

Light and flat field images are not corrected with the map of dark current, and isolated hot pixels are noticeable in these images. These artifacts are close to an uncorrelated signal, and are difficult to compress. These test images come from our DEIMOS [3] database, which is available as open source (<http://www.deimos-project.eu/>). This image database covers a broad range of image content from

scientific image data in astronomy and multimedia [6].

3 Lossless astronomical image compression

3.1 JPEG2000

The core part of the JPEG2000 standard [8] also enables lossless compression. For the lossless mode, the reversible color transformation and the reversible wavelet transform can be used to decorrelate the input data in terms of the color components and the spatial dependencies. These transformations convert input integer data into integer results. The reversible color transformation and the reversible 5/3 wavelet filter can also be used for lossy coding. Thanks to the sophisticated JPEG2000 format structure, it is then very simple to work with the quality or resolution progression, from a lossy image overview until lossless maximum resolution image data. The ROI (Region of Interest) technique also provides the most accurate data for the specific part of an image with reasonable bandwidth requirements. Although the compression performance of the reversible transformations is limited for the lossy case, they show results that are almost comparable with the irreversible transformations dedicated to lossy compression.

3.2 HCOMPRESS

HCOMPRESS was developed at the Space Telescope Science Institute (STScI, Baltimore), and is commonly used to distribute archived images from Digital Sky Survey DSS1 and DSS2. This compression format is based on the Haar transform (2×2 pixels). The computation is extremely fast, since the Haar transform does not require any multiplication. Wavelet coefficients are linearly quantized, quad tree coded on bitplanes, and the statistical redundancy is reduced by the Huffman code. Besides lossy coding, this compression format also enables lossless compression, since the Haar wavelet transform is reversible. A definition of this format can be found in [14].

3.3 CCSDS-LDC or Rice algorithm

The Consultative Committee for Space Data System published in 1997 a recommendation standard for lossless data compression based on modified Rice algorithm [15]. LDC stands for Lossless Data Compression. This coding should exhibit better results than JPEG-LS under the same conditions. The original Rice's algorithm can be found in [16].

3.4 ACC Coder

ACC stands for Astronomical Context Compression. This format is currently under development in the radio engineering department of FEE of CTU in Prague, and is being designed especially for astronomical images, focusing on their specific characteristics. However, it can also be applied for general raster images. ACC consists of the following main parts:

- background estimation
- successive spatial decomposition
- context computation based on noise evaluation
- context-based pixel estimation, using linear regression
- RLE and arithmetic coding

Background estimation is the first part of the coding process, and it is important. It is based on tiled median computation and subsequent filtering. The estimated background is extracted from the original image data, and the background-free image is further processed. This background separation improves the coding performance of the following methods.

The background-free image is then decomposed in several steps. In each step, a different set of pixels from the input pixel array is coded. Each pixel is coded just once, so the sets of the pixel are disjunctive. This spatial decomposition is optimized for the specific astronomical image data, where many singularities in the image function are expected. The decomposition scheme differs from the wavelet dyadic decomposition, where the input pixel array is processed in a successive pyramidal way.

Astronomical images are usually contaminated by a significant noise level. The key to the ACC algorithm is to measure the local noise characteristics and to differentiate the input image pixels into incompressible noise and significant data. According to the significance of the local data, the context is computed and assigned to each coded pixel. Pixels having the same context are then coded together.

4 Achievable compression ratios

The performance of the three compression formats presented above was measured and compared in terms of the lossless compression ratio that was achieved. The measurement was made on three image sets, each set representing a different astronomical image type. In the first set, there were 26 deep sky astronomical images. The second and the third set contained 22 correction dark frames and 5 correction flat fields, respectively. All tested files were 1536×1024 single component images with 16 bit/pixel depth.

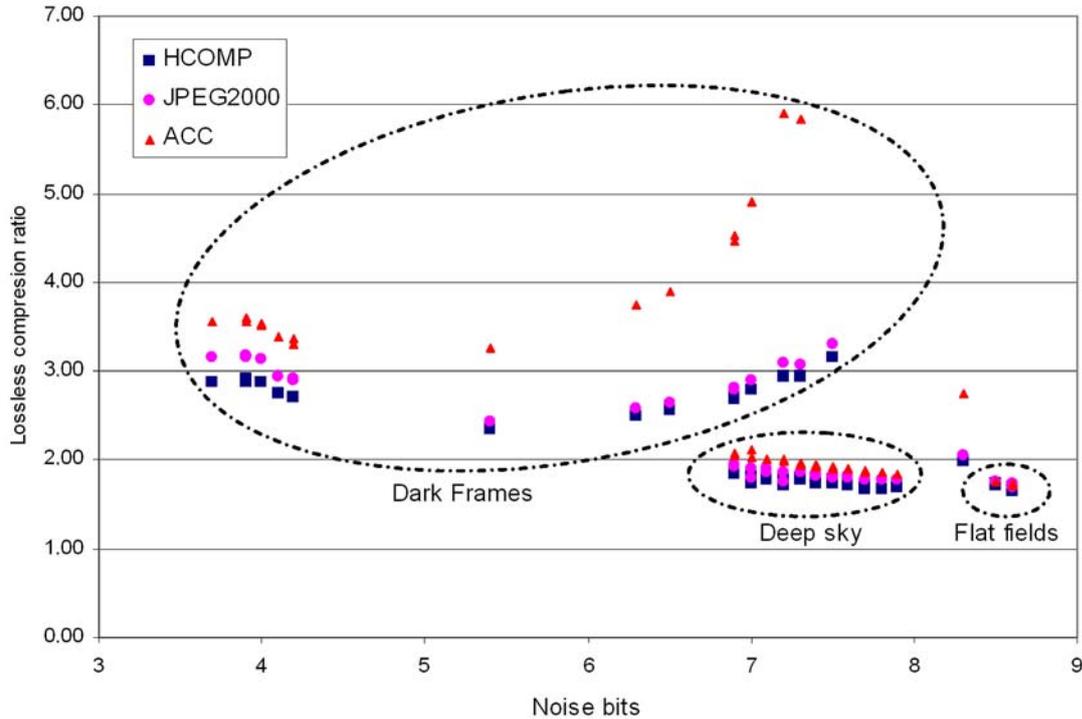


Fig. 3: Noise bits versus the lossless compression ratio

Table 1: Average lossless compression ratio

IMAGE SET	HCMOPRESS	JPEG2000	RICE	ACC
Deep sky	1.77	1.85	0.51	1.96
Dark frames	2.69	2.85	1.52	5.63
Flat field	1.71	1.74	0.49	1.74

Figure 3 shows the measured compression ratio versus the Gaussian noise equivalent bits. All image sets are included. The Gaussian noise bits are computed by the fpack utility [12]. Among the standard coders, the JPEG2000 standard achieved slightly better results than HCOMPRESS. Its main benefit is the MQ entropy coder. However, the static 5/3 DWT filter is not optimal in many cases. For example, the Haar wavelet used in HCOMPRESS produces less high amplitude coefficients in the case of isolated singularities, which are common in astronomical images.

The results show clearly that the ACC coder exhibits very good results on all tested images. It shows superior compression ratios in almost all test cases. The strength of the context-based estimation optimization can be exploited particularly in the dark frames test set, where the average improvement of this novel method compared to the other algorithms was particularly evident. The dark frames usually include much less Gaussian-like noise, and this enables it to have better theoretical compression ratios, e.g. compared with the deep sky images.

5 Lossy astronomical image compression

The algorithms required by astronomers are lossless. Their efficiency is limited. Unfortunately, they do not offer a higher compression ratio than 5 : 1 [10]. Is this enough? Inadequate results can be enhanced by the use of lossy algorithms. They provide a much better compression ratio, up to 100–200 : 1 for specific kinds of images. However, they also lead to increased errors in the reconstruction images. JPEG and JPEG2000 are the most widely known loss compression standards. They are preferred by graphics and web users. However, their usage for astronomical data compression is not optimal. These standards are optimized for human vision (i.e. perception based) and for so-called multimedia applications.

We are searching for optimal compression algorithms with the following characteristics

- highly efficient, with a good compression ratio
- lossless, or loss with known and optimized defect reconstruction

- a fast decompression algorithm — e.g. the coder and decoder of an archive machine can be non-symmetrical.

Scientific data is not processed by the human eye, but sophisticated algorithms are usually used. They are sensitive to other parameters than the eye. The mean square error is usually used for estimating the good quality of an approximated image signal. A special compression technique is therefore studied in this paper. We can compare it with algorithms based on the unique properties of wavelets and fractals as alternative coding methods [13]. The technique described in this paper has the lossy coder of the spectral coefficients of the Karhunen-Loève transform [11, 5]. It seems to be a better solution.

5.1 Distortion measurement of lossy coders

The measurement confirms the possibility of arranging the coder blocks to produce an accepted error and a sophisticated data stream. First, the most principal spectral components are important for a preview of the image and background function estimation, together with sensitivity correction. Next, the components can be used for searching objects and for high-precision astrometric and photometric measurements with a profile fitting. Suboptimal KLT decomposition has been found to be very suitable for astronomical data compression. The KLT coding of correction sensitivity images (so-called flat fields) can be performed up to 100 : 1, according to the image characteristics [9].

The light images are very well reconstructed for compression ratios about 30–60 : 1 (see Figure 4). A comparison of the impact of the wavelet, DCT and KL transforms on the deep sky is shown in Figure 5. The dark frames are a map of the thermally generated charge in the CCD structure. They are very difficult to code, due to their noise and their very stochastic character. Application of the designed KLT provides an insignificant result. The maximal accepted error of the reconstructed images

corresponds to a compress ratio of about 5 : 1. The lossless variant of the KLT coder is recommended for use for these images. Figure 4 shows a comparison of the mean error of the object position for the Karhunen-Loève coder and the adaptive wavelet transform, based on the JPEG 2000 standard.

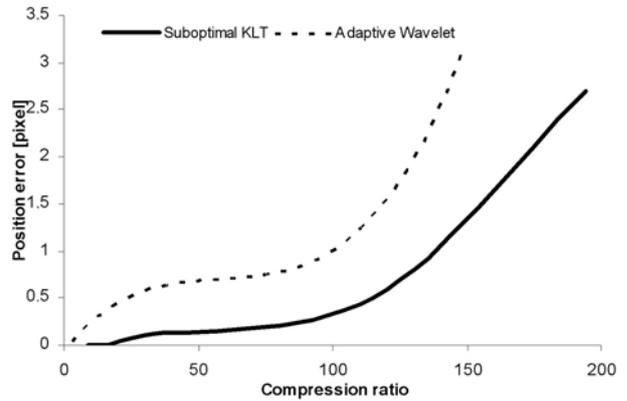


Fig. 4: Error of the astrometry position measurement for the suboptimal Karhunen-Loève expansion and the adaptive wavelet transform

6 Conclusion

The lossy compression technique described here can be considered as a good alternative for known compression algorithms (JPEG and JPEG2000). The disadvantages of the KLT-based coder are its extensive computational requirements due to the need to calculate the eigenvectors of the covariance matrix. It can be improved by using the suboptimal KLT coder. Further improvement of technique can be achieved by sophisticated filtering methods and suitable image data organization. The lossless ACC (Astronomical Context Coder) has been designed and optimized for specific astronomical data properties. The proposed new compression method is based on noise estimation and pixel contextual modelling using median regression. For a given context, this pixel estimation is optimal in the sense of the estimation error sum.

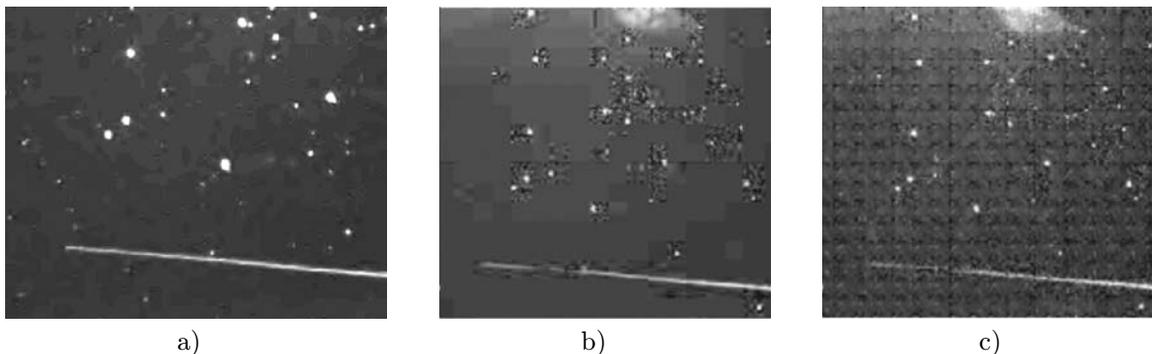


Fig. 5: Comparison of the impact of irrelevancy reduction for the Adaptive Wavelet Algorithm JPEG 2000 (a – left), DCT (JPEG) (b – central), and DKLT (c – right) based coder. Detail of object, stars and satellite tray in fig. 2b)

Acknowledgement

This work has been supported by grant No. P102/10/1320 “Research and modeling of advanced methods of image quality evaluation” of the Grant Agency of the Czech Republic, and by research project MSM 6840770014 “Research of perspective information and communication technologies” of MSMT of the Czech Republic.

References

- [1] Bernas, M., Páta, P., Hudec, R., Rezek, T.: Lossless and Lossy compression of Images from the OMC Experiment of Integral Project, *Astrophysical Letters and Communications*, Gordon and Breach Science Publishers, Amsterdam, 2000, 429–432.
- [2] Jelínek, M., Castro-Tirado, A. J., de Ugarte Postigo, A., Kubánek, P., Guziy, S., Gorosabel, J., Cunnife, R., Vítek, S., Reglero, V., Sabau-Graziati, L.: Four Years of Real-Time GRB Followup by BOOTES-1B, *Advances in Astronomy*, Vol. **2010**, 2010.
- [3] *DEIMOS*, Database of Images: Open Source, <http://www.deimos-project.eu/>.
- [4] de Ugarte Postigo, A., Mateo Sanguino, T. J., Castro Cerón, J. M., Páta, P., Bernas, M., et al.: Recent Developments in the BOOTES Experiment, In *AIP Conference Proceedings 662*. Cambridge : Massachusetts Institute of Technology, 2003, 553–555.
- [5] Effros, M., Feng, F., Zeger, K.: Suboptimality of the Karhunen-Loève transform for transform coding, *IEEE Transactions on Information Theory*, Vol. **50**, Aug. 2004.
- [6] Fliegel, K., Klíma, M., Páta, P.: New open source image database for testing and optimization of image processing algorithms, In *Optics, Photonics, and Digital Technologies for Multimedia Applications, SPIE Proceedings*, Vol. **7 723**, 2010.
- [7] Hudec, R., Soldán, J., Hudcová, V., Bernas, M., Páta, P., Hroch, F., Castro-Tirado, A. J., Mass-Hessem, J. M., Giminez, A.: *Blazar Monitoring towards the Third Millennium*, Torino : 1999, 131–133.
- [8] ISO/IEC 15444-1:2000: *JPEG2000 Image Coding System (core coding system)*, [online], 2000, <http://www.jpeg.org/FCD15444-1.htm>.
- [9] Páta, P.: Compression of Astronomical Images Based on the Karhunen-Loeve Transform. In *Proceedings of the Eighth IASTED International Conference on Signal and Image Processing*, Anaheim : ACTA Press, 2006, p. 133–138.
- [10] Páta, P., Bernas, M.: Properties of Karhunen-Loeve Expansion of Astronomical Images in Comparison with Other Integral Transforms, In *Gamma Ray Bursts*, AIP Conference Proceedings Woodbury : American Institute of Physics, 2000, 882–886.
- [11] Páta, P., Hanzlík, P., Schindler, J., Vítek, S.: Influence of Lossy Compression Techniques on Processing Precision of Astronomical Images, *6th IEEE ISSPIT conference*, Athens, Greece, 2005.
- [12] Pence, W. D., Seaman, R., White, R. L.: *Fpack FITS Image Compression Utility*, [online], 2010, <http://heasarc.gsfc.nasa.gov/fitsio/fpack/fpackguide.pdf>.
- [13] Starck, J. L., Murtagh, F., Louys, M.: *CCMA Conference on Data and Information Fusion*. Grenada, 1997.
- [14] White, R., Postman, M., Lattanzim, M.: *Digitized Optical Sky Survey*, Kluwer, pp. 167–175, 1992.
- [15] CCSDS: *Lossless Data Compression, Recommendation for space data system standards*, CCSDS, Vol. **121.0-B-1**, 1997.
- [16] Rice, R. F., Yeh, P.-S., Miller, W.: *Algorithms for a very high speed universal noiseless coding module*, JPL Publication 91-1, Jet Propulsion Laboratory, Pasadena, CA, 1991.

Jaromír Schindler
 Petr Páta
 Miloš Klíma
 Karel Fliegel
 Department of Radioengineering
 Faculty of Electrical Engineering Czech Technical University in Prague
 Technická 2, Prague, Czech Republic