Technical Report

Pattern Recognition and Image Processing Group
Institute of Computer Aided Automation
Vienna University of Technology
Favoritenstr. 9/183-2
A-1040 Vienna AUSTRIA
Phone: +43 (1) 58801-18366
Fax: +43 (1) 58801-18392
E-mail: nabil@prip.tuwien.ac.at
bischof@icg.tu-graz.ac.at
URL: http://www.prip.tuwien.ac.at/~nabil/pdf/

PRIP-TR-075

October 31, 2003

On Board Data Compression: Distortion and Complexity Related Aspects[1]

Ahmed Nabil Belbachir and Horst Bischof

Abstract

Digital signal Processing (DSP), and in particular data analysis and compression, has been studied for many years. However, only the recent advances in computing technology have made it possible to use DSP in day-to-day applications. Users expect the data to be transmitted in a minimum of time and to take up as little storage space as possible. These requirements call for efficient data compression algorithms in term of results quality and algorithmic complexity. The users want a good data quality with very fast compression and decompression as not to have to wait for data to be usable. Therefore, an accurate investigation on compression algorithms performances has to be performed. The performance of an algorithm can be analyzed using two criteria: the result quality i.e. high Signal-to-Noise Ratio vs Compression ratio and the algorithmic complexity. This report addresses both aspects of compression. In the first part of the report, noise influence on the compression method is stressed. The notion of algorithmic complexity is formalized using both approaches qualitative and quantitative in the second part of the report. In the third part, an new concept "Integrated data compression" is introduced. In this part, we provide an optimization of On-board compression algorithm in terms of algorithm complexity and results quality, introducing a distributed exploitation of the data in both sides remote and user side. Application results of the method for the case study “HERSCHEL/PACS Infrared Camera and Spectrometer” are presented in an other paper.

[1]This Work was supported by the Austrian Federal Ministry of Transport, Innovation and Technology.

Contents

List of Figures

Acronyms and Abbreviations

CR	Compression Ratio
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
FB	Filter Bank
FIR	Finite Inpulse Response
IR	InfraRed
IWT	Integer Wavelet Transform
KL	Kullback-Leibler
LS	Lifting Scheme
MAE	Mean Absolute Error
PDF	Probability Density Function
PMF	Probability Mass Function
PF	Polyphase Filter
PR	Perfect Reconstruction
PSNR	Peak Signal to Noise Ratio
R-D	Rate-Distortion
RMSE	Root Mean Square Error
SNR	Signal to Noise Ratio

1 Introduction and Motivation

Digital imaging has been a subject of research for several decades, but only lately has the technology revolutionized the consumer market. This explosion in the use of digital imagery can be attributed mainly to the significant advances in the processing power of computers. The recent development of the World-Wide-Web (WWW) has also played a key role in making image-data accessible over the Internet. One component that has lagged behind consumer-needs is the network-bandwidth. Satellite and video data imposes enormous storage and bandwidth requirements [5,42]. Therefore, although many applications exist that generate or manipulate these data, transmitting image information is still a significant bottleneck.
This constraint has stimulated advances in compression techniques for satellite data, still images and video data [3,4,5,42]. Although data compression is still a major topic of research, the industry has already come up with several compression standards, namely, ZIP, Real Networks, JPEG, JPEG2000, MPEG-1, MPEG-2 and MPEG-4 [8]. All compression techniques exploit the fact the data are highly redundant. This redundancy allows us to reconstruct a signal (e.g. image or video) from compressed information that is only a small fraction of the original data, in size. Of course, using compressed data implies that the consuming application must first reconstruct the signal information. Thus, we are in fact compensating for inadequate network bandwidth with on-board processor-power. As usually happens when technologies evolve, constraints on data quality grow with the challenging consumer demands and the algorithms for compressing and decompressing visual information is becoming increasing complex. In addition, the expectations of consumers for transparent and real-time decompression is also growing day by day. Therefore, significant research effort has also been invested in analyzing the performance of the compression algorithms in terms of results quality and complexity. Such an analysis forms the basis for optimizing the algorithms, and also for determining whether a given algorithm is appropriate for the application at hand.

In this report, we focus on data compression for infrared astronomy. Astronomy data collected on-board the space observatories impose enormous storage and bandwidth requirements for downlink [5,6]. Infrared detectors consist, as a rule, of less pixels but the design of multi-sensor instruments leads to even higher data volumes. If multiple detectors are operated in parallel to support multi-spectral or even hyper-spectral imaging, then, the data volumes multiply [40]. In most cases, the actual data rates to be sustained are dictated by the image repetition rate due to optimum footprint selection, maximum illumination levels, and avoidance of motion blur. Although many applications exist However, no real study has been performed for infrared astronomy data compression apart the use wavelet-based compression techniques [3,4]. The IR data per definition contains a high entropy due to the influence of noise [1,41]. In addition to that, the detectors are continuously exposed to high energy cosmic particles inducing a disturbance (glitches) of the readout voltage which decrease the signal to-noise ratio and hence the capability to achieve the required compression ratio. Therefore, dedicated compression technique has to be used in order to preserve the quality of the data.

In this report, some aspects of distortion/rate and complexity analysis for data compression algorithms are exposed. Although we will specifically consider only image and data compression algorithms for infrared astronomy, the techniques developed here can be used to analyze any signal processing algorithm. We will also present an optimized model for data compression that adapt the environment characteristics and better exploit the available resources.

This document can be subdivided into four main parts, the one dealing with the analysis of the performance of compression method in a noisy environment, and the second part introducing the notions "data compression" and "reconstruction error". In the third part, the compression algorithm challenges in term of complexity is depicted. An integrated data compression model is presented in the last Section. We conclude with a short summary

2 Preliminary Notions

This Section is subdivided into two parts. In the first part, we introduce the notion of data compression. We also briefly give some assessment criteria for the compression methods in the second part.

2.1 Definition of Data Compression

Data compression consists of finding and removing most, if not all, the redundancy from the data for an efficient data transmission or storage (Figure 1). The degree of data reduction obtained as a result of the compression process is known as the compression ratio. This ratio measures the quantity of compressed data in comparison with the quantity of original data [1].

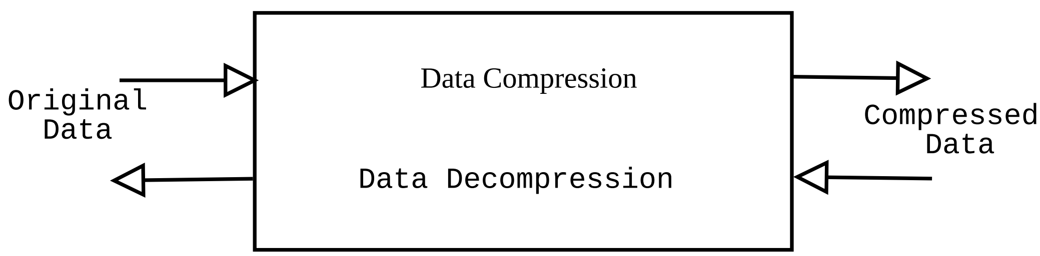

Figure 1: Basic Data Compression Block Diagram

Compression Ratio(CR) $= \frac{length-of-original-data-string}{length-of-compressed-data-string}$

Compression techniques can be classified into one of two general categories or methods - Lossless or Lossy- [1].

- Lossless compression techniques are fully reproducible and are primarily restricted to data operations. Data are compressed such their decompression should result in the exact reconstruction of the original data. Other common terms used to reference lossless compression include 'reversible' and non-destructive compression.

- Lossy compression techniques may or may not be fully reproducible and are primarily restricted to operations on images or on data sets. Although the result of decompression may not provide an exact duplication of the original data, the differences between the original and the reconstructed data may be so minor as to be difficult to see. Since the compression ratio obtainable by the use of lossy compression can significantly exceed the compression ratio obtainable from lossless compression, the primary trade off concerns the need for reproducibility versus the storage and transmission requirements.

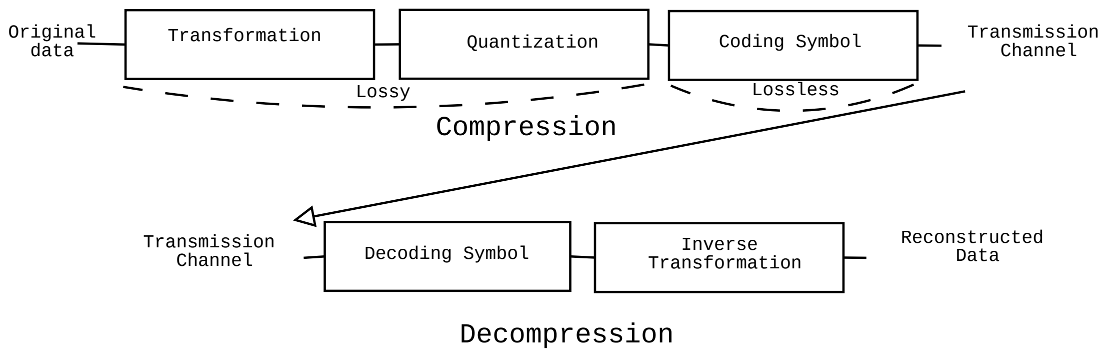

Figure 2: General Block Diagram for Data Compression/Decompression

The general block diagram for data compression/decompression is depicted in Figure 2 [1]. The lossy compression part consists of a decorrelation stage (usually DCT, Wavelet..etc) [,36] and a quantization stage. Lossless data compression consists of lossless entropy coding (e.g. Huffman, Lempel-Ziv, arithmetic coding...etc) [35]. The decompression part consists of reconstructing the original data by decoding and demultiplexing the compressed data stream. Constraints related to data compression are described in the following sections.

2.2 The Reconstruction Error

The reconstruction error is a parameter used to assess the performance of a compression method. It is applied to quantify the data distortion by means of the difference between the original and the reconstructed data. There are several quantitative measures for this distortion. These measures includes signal-to-noise ratio (SNR), root mean square error (RMSE), mean absolute error (MAE), peak signal-to-noise ratio (PSNR) and Kullback-Leibler information gain (KL gain); that is, the distortion measure often used in R-D (Rate-Distortion) optimization. In the following, we give the definitions of the measures listed above.

- MAE: This is the simplest measure. If S is the original data and S_R is the reconstructed data, then the mean absolute error is defined as follows:

$$MAE(S, S_R) = \frac{1}{N} \sum_{x=1}^{N} \mid S(x) - S_R(x) \mid \tag{1}$$

- RMSE: In contrast to MAE, the RMSE averages the squares of the differences and then the square root of the result is taken. The RMSE is an area weighted statistics. It is calculated as the standard deviation of all reconstructed data set relative to the original data. It defined as follows:

$$RMSE(S, S_R) = \sqrt{\frac{1}{N}\sum_{x=1}^{N}(S(x) - S_R(x))^2} \quad (2)$$

- SNR: The SNR attempts to improve the RMSE equation by taking into account the intensity of the reference. This is done by dividing the total reference power by the total error power. Taking the logarithm is one way of reducing the range of values, and the 10 in front is the convenience factor.

$$SNR(S, S_R) = 10log_{10}\frac{\sum_{x=1}^{N} S(x)^2}{\sum_{x=1}^{N}(S(x) - S_R(x))^2} \quad (3)$$

  Notice that this measure is inversely proportional to the previous two.

- PSNR: The PSNR criterion is widely used for the evaluation of compression methods. The reason is that PSNR is mathematically repeatable and is highly susceptible to different intensities.

$$PSNR(S, S_R) = 10log_{10}\frac{(max(S(x)) - min(S(x)))^2}{\frac{1}{N}\sum_{x=1}^{N}(S(x) - S_R(x))^2} \quad (4)$$

- KL gain: This measure, also known as a relative entropy, makes use of the probability distribution of the data. For instance, the original data S has a 'true' probability distribution Q and an estimated distribution E. The KL information gain (DG_{KL}) could be calculated as follows [33]:

$$DG_{KL}(Q, E) = a\sum_{x} p(x/S)log_2\frac{p(x/S)}{p(x/S_R)} \quad (5)$$

  where:
  - a≥0 constant,
  - p(x/S) is the probability of occurrence of the element x in the original data S and
  - p(x/S_R) is the probability of occurrence of the element x in the reconstructed data S_R.

  The KL information gain may be applied to to quantify the distortion by means of the difference between the original data and the data reconstructed after compression

The above listed metrics aim to measure the quality loss resulted from the applied data compression methods. However, It is not obvious to have a single metric that fit to the human interpretation (Human visual inspection). Our goal is to trace the curve (compression ratio vs. reconstruction error) resulting from the processing of selected data, combining the listed metrics, for an objective measure of the compression performance.

3 Compression Challenges

The performance of a data compression method can be evaluated using the following relevant parameters:

- The compression ratio vs the reconstruction error
- The complexity of the method
- The execution time

The first criterion generally used to assess the performances of a compression method is the compression ratio achieved. It points out to the capability of the method to find and remove the redundancy in the data. The more redundancy is removed the better compression rate is achieved.
The reconstruction error defines the quality of the data after reconstruction. The results quality criterion which can be retained for estimating the merits and performances of a compression method, in case of astronomy, falls under these heading:
+) Visual Aspect,
+) Signal to Noise Ratio,
+) Detection of real and faint objects,
+) Object morphology,
+) Astrometry,
+) Photometry and
+) Accuracy in localization.
Although the upper criteria are very important to design a compression method, the complexity of the algorithm is of bigger importance because it defines the feasibility of the method. This criterion (second criteria) has not to be performed at latest stage of the design, but it has to be taken as a criterion to assess a compression method. Therefore, the implementation of the method has to be part of the design of method.

The last criterion (Execution time) mainly depends on the algorithm implementation and on the machine where it is running. The relationship between the algorithmic and the execution time is very strong but it is not trivial if the way the method is implemented and the hardware description are not known.

As described in Figure 2, data compression consists of two steps: lossy and lossless. Generally, lossless compression leads to low data compression factor(up to 5 in case of astronomy) especially, for noisy data where the degree of randomness is very high. Therefore, lossy compression is mandatory for several applications (radar, space applications, telecommunications) [36] to achieve additional compression ratio. Although certain operational applications (e.g. CCD images of deployment of solar panels or position of landing gear) allow a high amount of quality loss, noisy data has to be compressed with care that real objects (e.g. faint objects) are not lost while compression.

One of the most common methods for lossy compression(Figure 2) is transform coding e.g. DCT [18], wavelets [12], curvelets [39], ridgelets [39]... etc. The objective is to minimize statistical dependence between the output values of the transform. Desirable characteristics are decorrelation, energy concentration, and computational efficiency. All these transform assume that the original data is free of noise or that noise is a part of significant data elements. In addition to the quality loss due to artifacts (blockiness) problem [1,8,36] which limits the capabilities of these transforms, significant data (e.g. faint objects in images) may not be reconstructed which leads to information loss. This is because some information with equivalent amplitude to the noise has been neglected.
Therefore, more noise models have to be considered before compression in order to preserve as much as possible the quality in the data. In this report, we present a new model which takes into consideration several models of noise before the data compression such that the significant data element even with low amplitude are reconstructed. One problem is that the more noise models we have, the more complex is the method. We discuss in the next Sections about the distortion and complexity aspects for compression purpose.

4 Noise Study

Data and images generally contain noise. In most applications, it is necessary to know if data element is due to signal (i.e. it is significant) or to noise. Generally, noise in astronomical data follows a Gaussian or a Poisson distribution, or a combination of both, if we only consider the detectors noise, the amplifier noise, the electric cross-talk noise or the pick-up noise [8]. Furthermore, astronomical data suffer from the cosmic ray impacts (glitch) and the transient behavior of the detectors, which may cause a potentially scientific data loss in case of integrating over several samples for compression purpose. Therefore, a deep investigation has to be made in order to prevent this noise before data compression especially, if we know that infrared space observatories are frequently confronted with this type of noise. In the following, we summarize the existing noise models and presents the challenge of data compression algorithms for noisy data.

4.1 Thermal or Johnson Noise

Johnson and Nyquist in the 1920s studied the noise resulting from the thermal agitation. This thermally generated noise produces a spectrum that has about the same energy for each cycle of bandwidth. This equal-power per cycle noise is termed "Gaussian" or "white noise". In the other hand, we can have shot of Gaussian noise if the electric current (change flow) does not flow in a uniform, well-behaved manner. Shot noise is often termed "Rain on the Roof" noise. This noise is usually quantified using the probability and statistics. This noise follows the Gaussian distribution with zero-mean and standard deviation σ. There are different ways to estimate the standard deviation of Gaussian noise in an image [9]. Olson [9] made an evaluation of six methods and showed that the average method was best, and this is also the simplest method. This method consists of filtering the image I with the average filter and subtracting the filtered image from I. Then a measure of the noise at each pixel is computed. Equation 6 from [44] gives the mathematical form of the Johnson noise V_n in voltage.

$$V_n = 2\sqrt{KTRB} \tag{6}$$

where:
- V_n is the noise voltage (V/$\sqrt{Hz}$)
- K is Boltzmann's constant (1.38 X 1023 J/K)
- T is the temperature in degrees Kelvin (K)
- R is the resistance in ohms (W)
- B is the bandwidth in hertz (Hz)

4.2 Flicker and 1/f Noise

In addition to the fundamental Johnson noise, many devices exhibit a second noise phenomenon caused by the flow of electric current. Electron or charge flow (current) is not continuous, well-behaved process. There is a randomness that produces an alternating current (random AC) on the top of the main direct current flow. This noise is difficult to quantify and to measure. This noise, in general, has a 1/f spectrum and is termed ”pink noise”. Pink noise has equal power per octave of frequency (log2 scale). An empirical description of the 1/f noise after Hooge [45,46] is a spectrum with

$$C_{1/f} = \frac{\alpha}{N_{tot}}.\frac{1}{f} \tag{7}$$

where N_{tot} means the total number of moving charges in the device and f is the frequency. The Hooge-Parameter α is a material characteristic.

4.3 Poisson Noise

In addition to the mean and variance, we can also discuss noise in terms of the shape of the distribution for each data element. One common distribution for the values of each element is determined by the nature of light itself. Light isn't a continuous quantity, but occurs in discrete photons. These photons don't arrive in a steady stream, but sometimes vary over time. Think of it like a flow of cars on a road–sometimes they bunch together, sometimes they spread out, but in general there's an overall average flow. Discrete arrivals over a period of time are modeled statistically by a Poisson distribution. If the noise in the data I is Poisson, the Anscombe transform [8]

$$t(s(x)) = 2\sqrt{s(x) + \frac{3}{8}} \tag{8}$$

acts as if the data arose from a Gaussian white noise model (Anscombe, 1948), with σ=1, under the assumption that the mean value of s is large.

A Poisson distribution [47] is described by Equation 9

$$P(i) = \frac{m^i}{i!}exp(-m) \tag{9}$$

where m is the mean and i!=1x2x3x...xi. In Poisson distribution the mean=variance. It is similar to a Gaussian distribution with the following exceptions/properties [8]:

- A Poisson distribution is for discrete values, not continuous ones.
- A Poisson distribution applies only to non-negative quantities–one counts arrivals, not departures.
- A Poisson distribution has the property that its variance is equal to its mean.

4.4 Gaussian and Poisson Noise

For small Poisson parameter values, the Anscombe transformation looses control over the bias [8]. Small numbers of detector counts will most likely be associated with the image background [8].i.e. errors related to small values carry the risk of removing real objects, but not of amplifying noise, because at increasingly low values, the pixel value is increasingly underestimated. Therefore, an extension of the Anscombe transformation has been performed by Bijaoui [34] to cope with the Poisson model limitation by taking the combined noise into account.

The arrival of photons, and their expression by electron counts, on CCD detectors may be modeled by a Poisson distribution. In addition, there is an additive Gaussian readout noise. Consider the signal s(x), as a sum of a Gaussian Variable, γ, of mean g and standard deviation σ; and a Poisson variable, n, of mean m_0. We set s(x)=γ + αn. where α is the gain.

The generalization of the variance stabilizing Anscombe formula is:

$$t = \frac{2}{\alpha}\sqrt{\alpha s(x) + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g} \tag{10}$$

With appropriate values of α, σ and g, this reduces to Anscombes transformation (Equation 8).

4.5 Other Types of Noise

The types of noise considered so far correspond to the general cases in astronomical imagery [8]. We describe now briefly methods which can be used for non-uniform and multiplicative noise.

- Additive non-uniform noise: If the noise is additive but non-uniform, we cannot estimate the standard deviation for the whole data. However, we can often assume that the noise is locally Gaussian and we can compute a local standard deviation of the noise for every data element. In this way, we obtain a standard deviation map of the noise, $s_\sigma(x)$).

- Multiplicative noise: If the noise is multiplicative, the data can be transformed to a logarithmic scale. In the resulting signal, the noise is additive, and a hypothesis of Gaussian noise can be used in order to find the detection level at each scale.

- Multiplicative non-uniform noise: In this case, we take the logarithm of the data, and the resulting signal is precessed as for additive non-uniform noise above.

- Unknown noise: If the noise does not follow any known distribution, sigma-kappa clipping method [35] can be used for the reduction of the noise in the data.

4.6 Rate-Distortion Principle

Lossy compression is a typical engineering trade-off: lower data quality for higher transmission rate. If the rate decreases a large amount for a small decrease in data quality, then lossy compression is usually considered desirable. For this reason, lossy compression is often evaluated with reference to a rate-distortion curve.

The relationship between the rate and the distortion in a signal is depicted in Figure 3. Note that R (entropy) unit is bits/sample while the D (distortion) represents the reconstruction error.

It is shown that a lower bit rate R allows some acceptable distortion D in the signal. In Figure 3(b) and (c), the equivalent constrained optimization problems , often unwieldy, are given. Minimization of one parameter (R or D) is only done for a given reference (D or R).

There exist several formulation of the R-D problem for instance:

- Shannon Source Coding Theorem and Converse [37]: For a given maximum average distortion D, the rate distortion function R(D) is the achievable lower bound for the transmission bit-rate. R(D) is continuous, monotonically decreasing for $R > 0$ and convex. Equivalently, we can use the distortion-rate function D(R).
- The Lagrangian formulation [38]: Instead of the cost function D, with constrained R, or R, with constrained D, we use the unconstrained Lagrangian cost function for a convex R-D function and non-increasing, for $D > 0$, subject to minimize

$$\mathrm{J} = \mathrm{D} + \lambda \mathrm{R}$$

- The Blahut's algorithm [38]: This algorithm intends to solve the Langrangian- formulation of the R-D problem. It is used to compute the R-D bound or the optimality testing by applying it to the tentative testing solution.

  For fixed probability mass function (PMF) $q(y|x)$, optimal PMF q(y) is:

$$q(y) = \sum_{x \in X} q(y|x)p(x) = q(y).f(y) \tag{11}$$

  where p(x) is the probability density function (PDF) of x.

  For fixed q(y), optimal $q(y|x)$ is:

$$q(y|x) = \frac{q(y)e^{-d(x,z)}}{\sum_{z \in Y} q(z)e^{-d(x,z)}} \tag{12}$$

  where p(y) is PDF of y and d(x,z) is the distortion parameter for (x,z).

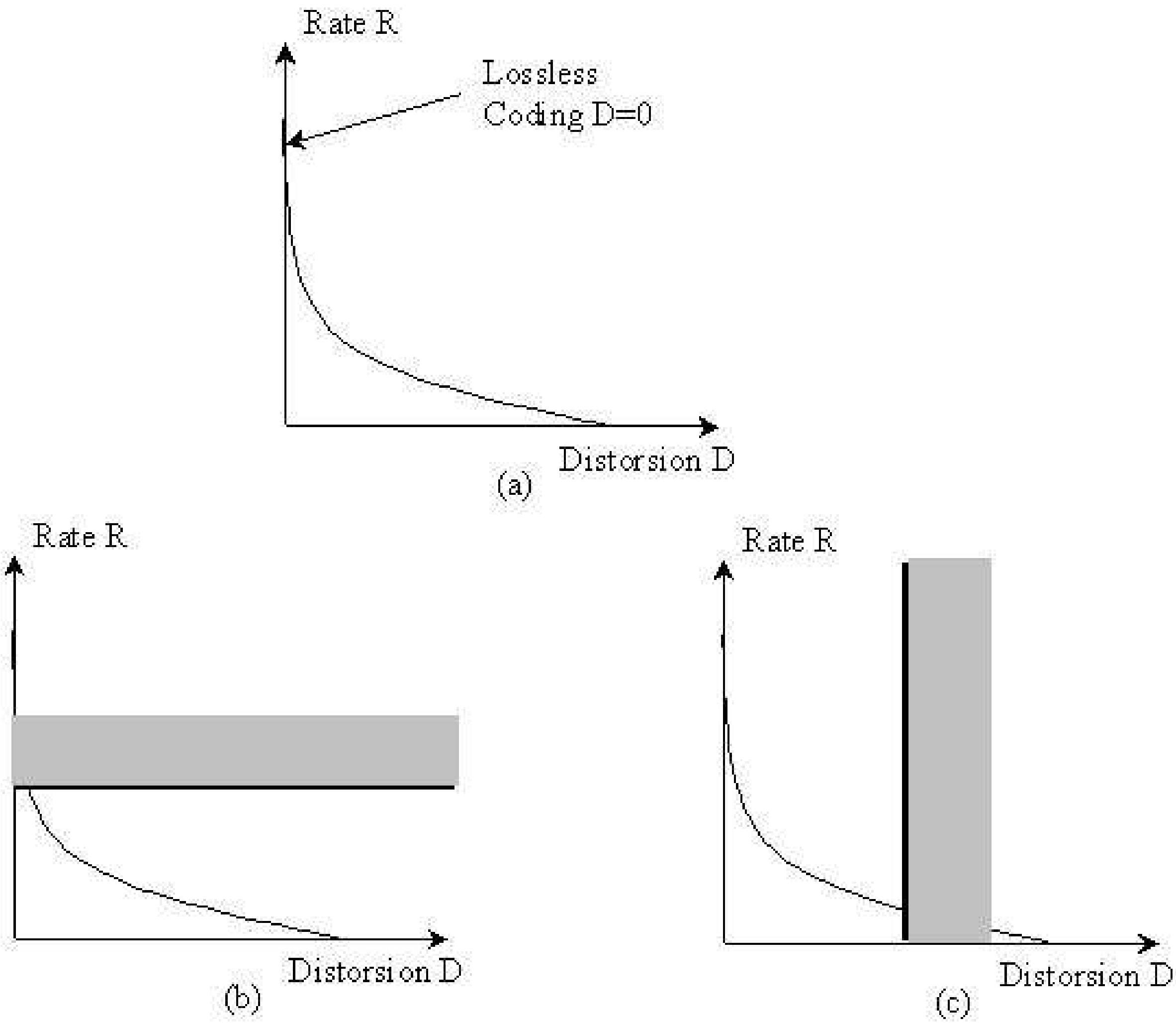

Figure 3: (a) Compression Rate and Distortion Trade off. (b) and (c) Equivalent Constrained Optimization Problems giving a Maximum Rate or Distortion

- The Kuhn-Tucker optimality conditions [38] for optimum q(y):
  $f(y) = 1$ if $q(y) > 0$
  $f(y) \leq 1$ if $q(y) = 0$

  This method set several constraints on R and D in order to investigate on the problem of optimality of the solution.

5 Complexity Definition and Measures

5.1 Definition

The complexity of a given algorithm can be analyzed in two ways: qualitatively and quantitatively [10]. Qualitative analysis usually means determining the complexity of the algorithm relative to other algorithms. Such analysis does not usually allow us to quantify the difference in complexities between algorithms. Quantitative analysis methods assess the complexity of a given algorithm in absolute terms, and often provide a numerical estimate of the complexity. Therefore, using quantitative analysis, it is also possible to state, for example, that algorithm A is three times less complex than algorithm B. Currently very few methods exist to measure the complexity of entire signal processing algorithms. In the case of simple algorithms, such as the Discrete Cosine Transform (DCT), the number of arithmetic operations can give a good estimate of the algorithm complexity. However this is not sufficient for more complex schemes such as image compression algorithms.

5.1.1 Qualitative Complexity

The first approach taken to understand and optimize the complexity of signal processing algorithms is based on an intuitive understanding of the complexity. For instance let us say that it is possible to find two algorithms realizing the same task. Suppose, the first one is based on integer arithmetic, whereas the second one uses floating-point arithmetic. We know that even with the most modern computer architectures, floating-point operations are more complex than their integer equivalents. However, it is difficult to quantify this difference without considering a specific hardware architecture. As another example, suppose we know that two algorithms differ in complexity because of the amounts of memory they require. However the relationship between the size of the memory and the complexity is not defined. Therefore, the difference in complexity is not quantifiable. Nowadays, the wavelet transform, or more precisely, the Discrete Wavelet Transform (DWT), is one of the most commonly used signal processing tools. It is very demanding in terms of arithmetic and memory operations. There exist fast wavelet-transform algorithms based on filter banks [11,12], and the lifting scheme [13,14]. The latter technique allows us to compute an integer version of the DWT, called the Integer Wavelet Transform (IWT). Evidently, the IWT has lower complexity than the original DWT. It can also be used in wavelet-transform based algorithms for lossless image compression.

5.1.2 Quantitative Complexity

Qualitative analysis gives us only an intuitive understanding of the complexity of an algorithm. It is more interesting to be able to quantify the reduction in complexity as a result of some optimization. A measure of the complexity is also needed to compare different algorithms realizing the same task.

Image compression is a typical example. Several algorithms can compress a given image with approximately the same efficiency. How can we compare these algorithms? The nature of the algorithms and the reasons for their compression-efficiency may be based on very different principles. Moreover the algorithms cannot be characterized purely in terms of arithmetic operations. This makes it difficult to use common measures of the complexity. This leads us to consider new methodologies to measure the complexity of these algorithms. Such a measure should not only incorporate notions linked to the number of arithmetic operations, but also the number of memory operations. Another aspect that should also be taken into consideration is the number of logical tests, or branches. A realistic measure should incorporate all those aspects. In practice it is very difficult to apply such a comprehensive complexity-measure to an entire algorithm as a unit. However, most algorithms can be broken down into smaller components. For instance, compression algorithms are often based on a transform block followed by a quantization step, and finally an entropy coding scheme. Many compression algorithms use the same transform and the same quantization procedures. Therefore, the complexity of these blocks need be computed only once.

Finally, as mentioned before, a quantitative measure of complexity can also be used to compare algorithms. This is especially appropriate for comparing data compression algorithms which vary greatly in nature, and therefore cannot be easily compared in qualitative terms. Furthermore, quantitative complexity-analysis can be effectively used to design rate-complexity or distortion-complexity functions for compression algorithms. This is a logical extension of the common rate-distortion approach used in these algorithms.

5.2 Software Complexity

One way to measure the complexity of signal processing algorithms is to look at their resulting software implementations. The complexity of the software can be measured. This gives information concerning the original algorithm. This solution is only an approximation as the notion of complexity is often very different between the programming and the signal processing worlds. In fact, the most commonly used complexity measure of signal processing algorithms is the execution time on a given implementation. A review of many software complexity measures can be found in [15]. The different methods are compared, pointing out the aspects they can handle and the ones they fail to incorporate. However, it is possible to select a subset of measures that are similar to the aspects we want to measure.

5.3 Implementation Complexity

One of the major approaches to the construction of correct concurrent programs is successive refinement: start with a high-level specification, and construct a series of programs, each of which "refines" the previous one in some way. In the realm of shared-memory concurrent programs, this refinement usually takes the form of reducing the grain of atomicity of the operations used for interprocess communication and synchronization. For example, a high-level design might assume that the entire global state can be read and updated in a single atomic transition, whilst a low-level implementation would be restricted to the operations typically available in hardware: atomic reads and writes of registers, test-and-set of a single bit, load-linked/store-conditional, compare-and-swap, etc. Each of the successive refinements is considered correct if and only if it conforms to the specification. The notions of conformance to a specification leads to the implementability concept or, in other terms, how challenging is the software implementation. In this work, we are concerned of assessing the implementability of the compression concept, we develop, for IR astronomy on DSP.

5.4 Complexity Measurements and Optimizations

Some signal processing algorithms have already been studied in terms of complexity. Often, only the pure arithmetic complexity (additions, multiplications) was studied. Moreover, the studied algorithms correspond generally to the transform parts of the codec and are thus mostly arithmetic based. One of the pioneers in the domain is Winograd, who studied the multiplicative complexity of many basic signal processing tools. He started with Finite Impulse Response (FIR) filter analysis in [16]. He then analyzed the Discrete Fourier Transform DFT [17] and the DCT [18]. Most of the time, the analysis is followed by an optimization of the scheme according to the measured features. All the studied algorithms are perfectly deterministic and the complexity is independent of the input data. Another basic signal processing block that has been often studied is the Fast Fourier Transform (FFT). In [19] different FFT algorithms are compared in terms of computational speed, memory requirements, implementation complexity, ease of testing and accuracy. However, the FFT is rarely used in data compression, because it introduces complex numbers. It's real version, i.e. the Discrete Cosine Transform "DCT" is generally used instead. Fast algorithms have been designed where the number of multiplications is minimized [20]. The Discrete Wavelet Transform "DWT" was also studied in terms of arithmetic operations in [21]. However the analysis of the transform is very rudimentary. A better comparison of many implementations of the transform can be found in [22].
In the framework of the JPEG algorithm, many data dependent complexity measurements and optimizations have been performed. Each time, only the DCT part of the algorithm has been studied. Variable complexity inverse DCT algorithms were proposed in [23] and [24]. In both cases, the algorithm uses the fact that the DCT coefficients are mostly zero because of the quantization. The same type of optimization was also proposed for the forward DCT [25] and the inverse DWT [26]. The complexity is measured using arithmetic operations, branches and memory accesses. The number of operations are summed after having being multiplied by the following weights: 1 for additions and shift, 3 for multiplications, 5-6 for tests and 1 for each memory access. All those algorithms lead to the same output as the conventional version of the transform. There is only vague information about how those weights were determined.
Another approach to complexity optimization is to trade complexity for the accuracy of the transform: the complexity can be reduced if the algorithm is allowed to give only an approximation of the correct transform. This strategy has been used for the DCT transform in [27]. This simplified inverse DCT sets some coefficients to zero in order to reduce the number of operations. The images are compressed using the JPEG codec at different

bitrates. The total distortion, due to the approximate DCT and to the compression, is measured as a function of the number of multiplications. This allows plotting the complexity/distortion curve of the algorithm. Varying the size of the block can also be used to modify the relationship between the complexity and the performance of the transform. Let us recall that the DCT is conventionally performed on blocks of 8x8. In [28], the block size of the transform is varied in order to find a R-D optimum. The number of multiplications is used to measure the complexity. The rate and the distortion are predicted analytically using a first order regressive model and a Laplacian distribution. Even if the primary aim of the author is to find a R-D optimum, a 3-D plot of the complexity/distortion/rate function is provided.
Much less research was done concerning memory utilization. However, JPEG is already relatively memory efficient because of the block structure of the DCT. On the other hand, DWT based algorithms do not naturally use a block structure. This has encouraged algorithms minimizing the overall memory demands. The memory bandwidth of an optimized DWT codec was studied in [29]. This algorithm corresponds to one mode of JPEG2000.
Again, most of the previously cited works concentrated on the transform part of the algorithm. To my knowledge, only Nielsen [30] and Reichel [10] have tried to measure the complexity of a full compression system. Nielson developed a model for the complexity and applied it to voice coding algorithms. Sadly, the paper is incomplete and this approach was not followed in other papers. The measure was based on arithmetic and memory complexity (reads and writes), "smartness of the programmer", parallelism and architecture considerations. A simple example of FIR filtering is used to demonstrate the technique. The following articles on the subject did not investigate the problem any further and use only execution time as a measure of the complexity [31]. Reichel [10] studied the complexity related aspects for image compression algorithms. Unfortunately this study only concentrates on the Wavelet-based compression methods and therefore could not be extended to a wide range of compression methods.

The aim of this work is to investigate new measure of the complexity of data processing algorithms. The proposed measure takes into account arithmetic operations, tests (or branches) and memory operations. The measure works in two steps, one depending on the algorithm, and the other, on the architecture on which the algorithm is implemented. The complexity of the algorithm is then expressed by a weighted sum of the algorithm-dependent counters, where the weights are determined by the architecture-dependent step. However, even with this well-defined methodology, analysis of the complexity analysis is still a long and difficult process. One way to simplify the problem is to use the fact that most algorithms can be divided into a succession of small tasks (or blocks). This is especially true for image compression algorithms. Therefore, complexities of the most common processing blocks for image compression are studied separately. A New algorithm complexity analysis becomes, then, a sum of the complexity of each one of its building blocks. The global aim of this work is to develop, improve and optimize a selection of tools and methods for data compression of infrared astronomical data. It is clear that infrared space astronomy imposes specific requirements, which consist of fulfilling the bandwidth limited downlink requirement for a minimal data loss by means of limited on-board resources (memory and CPU power) due to the limited budget of the spacecraft and the increasing cost of space-qualified devices. Nevertheless, the majority of well-established data compression techniques are increasingly complex [48,49] and thus rapidly overload the on-board resources. Therefore, we developed and new concept to improve the reliability of transmissions by considering on-board integration and data caching and by developing new algorithms, which adapt the compression ratio to the available bandwidth, and by defining optimal schemes for coding information for different telemetry rates. This topic is essential since it would lead to effective use of bandwidth and multiplicity of astronomical products. These developments are then applied to IR astronomy where data have high entropy due to the influence of noise [7,40].

6 Integrated Data Compression

A classical concept for compression, transmission and processing is reported in Figure 4-a. Such a scheme is becoming inadequate for airborne or spaceborne missions, where the information rate largely exceeds channel capacity and high compression rate is mandatory to fulfill the transmission requirements. In addition to that, the detectors are continuously exposed to high energy cosmic particles inducing a disturbance (glitches) of the detector signal and transient behavior of the detector which may increase the degree of randomness in the signal data and hence the capability to achieve the required compression rate even with the most powerful compression algorithms.

Therefore, it is more convenient that some processing is performed on-board i.e. to move a part of data processing from the user side to the remote side. In other words, the alternative to a robust data compression is to make the on-board sensor intelligent, and hence capable for interpreting the collected data. In the scientific literature several data compression algorithms exist, which are nearly optimal for a wide class of data. They make use of information extraction methods in order to reduce the original data, by keeping the relevant information (parameters). On the other hand, it turns out that techniques for information extraction strongly depend on the kind of information or pattern to be sought and recognized. The most general techniques are based on template matching, but they are also computationally expensive; more effective methods are based on detection-classification in transformed domains, and require to focus on a single or on a very restricted class of templates.State-of-the-art data compression algorithms usually employ transform coding in order to achieve a decorrelated signal which is more suitable for efficient coding. As already stated, pattern detection and classification can be carried on in a transformed domain as well; clearly, in order to ensure the maximum efficiency, a compression and detection algorithm should be devised, where the two operations are performed in the same transformed domain.

The global aim of this work is to develop, improve and optimize a selection of tools and methods for data compression of infrared astronomical data. It is clear that infrared space astronomy imposes specific requirements, which consist of fulfilling the bandwidth limited downlink requirement for a minimal data loss by means of limited on-board resources (memory and CPU power) due to the limited budget of the spacecraft and the increasing cost of space-qualified devices. Nevertheless, the majority of well-established data compression techniques are increasingly complex [48,49] and thus rapidly overload the on-board resources. Therefore, we developed and new concept to improve the reliability of transmissions by considering on-board integration and data caching and by developing new algorithms, which adapt the compression ratio to the available bandwidth, and by defining optimal schemes for coding information for different telemetry rates. This topic is essential since it would lead to effective use of bandwidth and multiplicity of astronomical products. These developments are then applied to IR astronomy where data have high entropy due to the influence of noise [7,40].

The new concept is presented (Figure 4-b) in the case where the patterns of interest are linear features. It uses a distributed processing on the remote and the user sides,

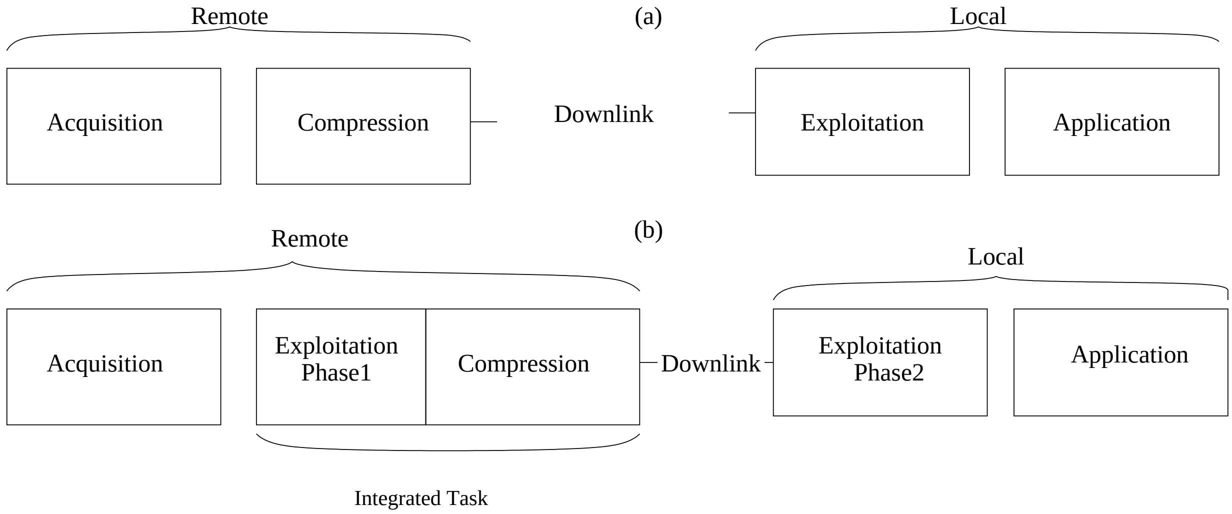

Figure 4: a.) Classical Data Compression b.)Data Compression Model using Distributed Data Exploitation "Integrated Processing".

so imposing to resort to lossy data compression, with potentially negative effect to the applicative tasks to be run at the ground station. A model-based approach is used at the remote side in order to reduce the data according to the model available on-board. This concept has been tested for infrared astronomy where different data models and noise models are available in order to support the on-board data compression. These models have been developed with care taking into consideration the infrared signal characteristics [40] and the ISO experience with noise and heavy ion and electron impacts [7]. Besides allowing for data reduction at the airborne segment by means of a simple detection algorithm, this would yield a strongly reduced computational burden, and the capability of performing a refined detection stage at the ground station on the decompressed data. Furthermore, knowing the data model yields to a very high compression efficiency (CR).

The presented model is subdivided into two parts:

Remote Side: On-Board Processing

It consists of 3 steps:

1. Acquisition: Acquisition systems transform variations in object radiant energy into an image or an electrical signal from which an image can be reconstructed. Images formed by optical system that projects radiation into a photo-sensor in image plane. The photo-sensor converts the radiation into a latent image into an electric signal that is amplified, sampled and quantized for digital transmission.

2. Exploitation Phase1: Figure 5 shows roughly the steps performed within this phase. The basic idea is that in order to achieve the high compression ratio we have to do on-board processing. First, we have to ensure that the noise level in the data is very

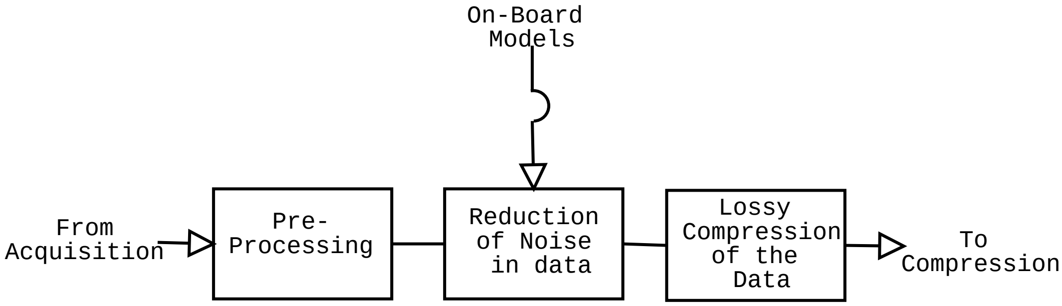

Figure 5: Model-based On Board- Data Reduction- Exploitation Phase1-

low such that the reduction does not affect the significant information in the data. Therefore, we need different parametrized models of noise. Then, we compute a set of parameters from the input data in order to match the current entropy model in the data with one of the noise models. Hence, the noise (considered as non-relevant information) is lowered from the data before the data reduction. In the case of infrared astronomy, we are using resampling algorithm for data reduction i.e. data is downsampled to a predefined rate to achieve an acceptable compression ratio. We can notice that the reconstruction error of the data mainly depends from the robustness of the modelization and the glitch removal.

3. Compression: This part consists of the entropy coding of the resulted data. It consists of the reduction of the redundancy (temporal, spatial and/or frequential). Backend entropy encoding algorithm (arithmetic coding) is also used to reduce the statistical redundancy. Again, the efficiency of lossless compression depends on the performane of the noise modellization i.e.the less entropy in the data, the more performant in the compression method

User Side: On-Ground Processing

It consists of 2 steps:

1. Exploitation Phase2: This part is performed to restore the data and prepare it for the application, predicting the negative effect of the tasks of exploitation phase 1. Indeed, the accuracy of the on-board modeling of the noise. A multiscale filtering method [8] based on the Wavelet transform is used for the restoration of the data.

2. Application: During this steps the reconstructed data/images could be used for the purpose they were intended for e.g. stars follow-up, investigation on water on planets, estimation of biomass in a region...etc

A detailed experimental evaluation of this model on astronomical data could be found in another technical report, where this concept has been tested for the 6-PACK data from HERSCHEL/PACS photoconductor camera [32].

7 Illustration of IR Data

The aboce-described concept will be evaluated on IR data from ISOCAM [50] and HERSCHEL/PACS [51]. Figure 6 depicts three image models from ISOCAM. One can clearly see the black column that is a consequence of pixels defection in-orbit. Figure 7 depicts a 1D signal from 8 different PACS detectors for a time constant of 0.25 seconds.

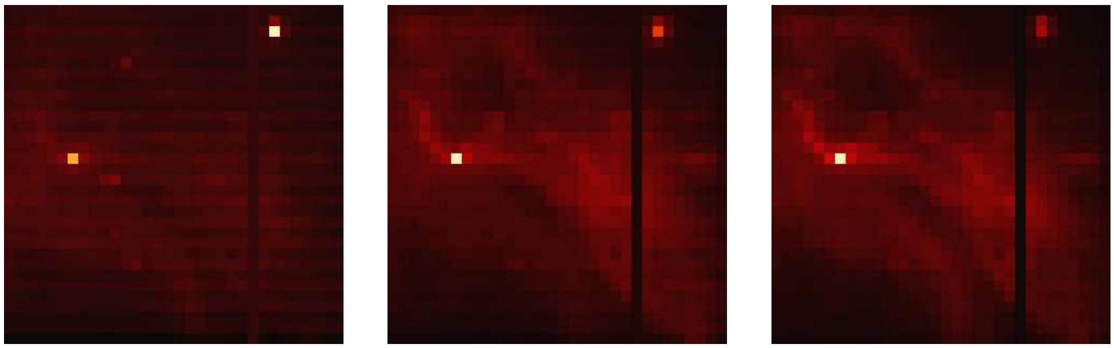

Figure 6: Selected ISOCAM Images

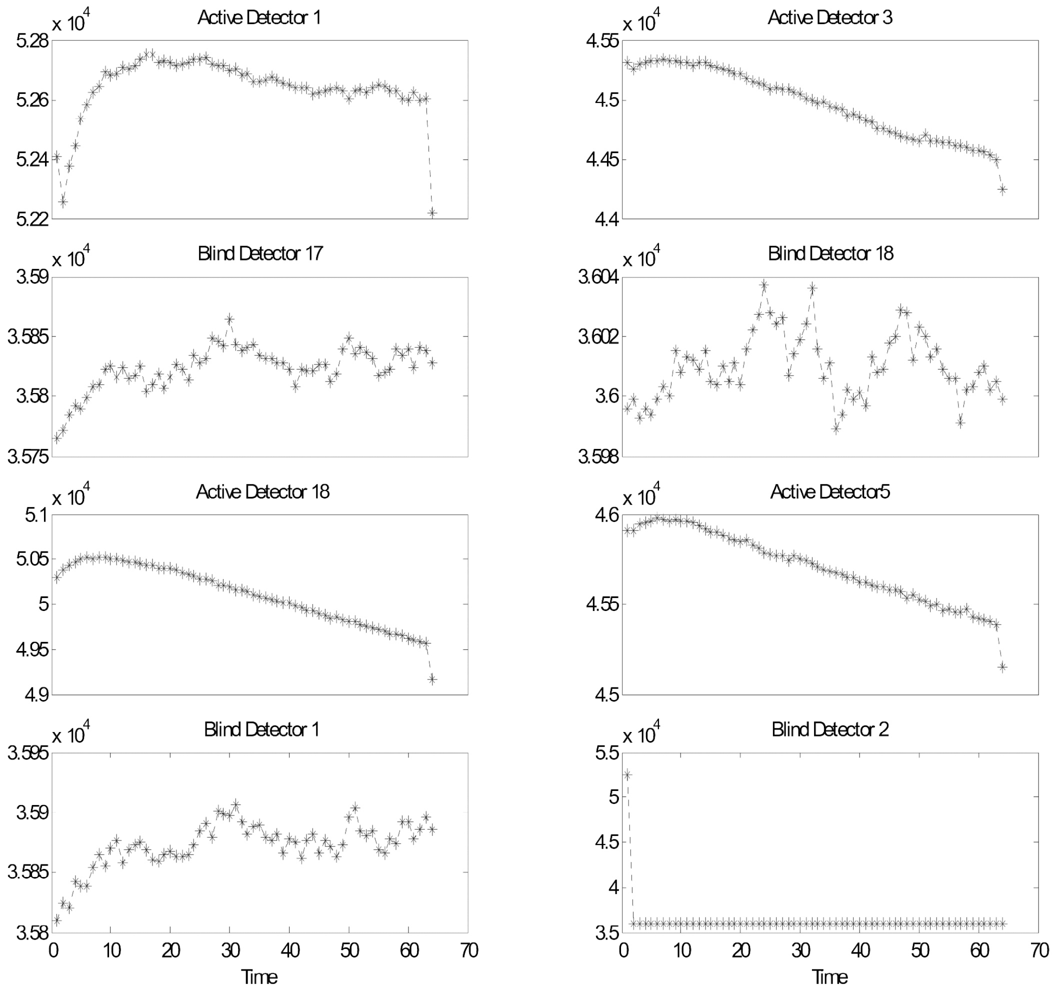

Figure 7: PACS Data from Selected Detectors for 0.25 sec Time Constant

8 Conclusion

In this report, we studied the performance aspects of compression methods in term of result quality and algorithmic complexity. As noise is by definition non compressible, a big challenge of the most recent compression method in a noisy environment is to achieve the required compression rate with a minimal information loss. We present different models of the noise and different approaches handling rate-distorsion problem. Additionally to compression performance and quality, we stressed the importance of algorithm complexity for the evaluation of a compression method. Indeed, the complexity of an algorithm defines the feasibility and the applicability of the method. Furthermore, existing methods for computing the complexities of algorithms were briefly introduced. Following that, two approaches for complexity-analysis were presented. The first one analyzes and optimizes the complexity in an intuitive manner. The second approach proposes a methodology for measuring and quantifying the complexity.
In this report, we present the challenges of compression method for space application. We stress the influence of the noise for the performances of compression methods. Finally, we present a model for data compression/Decompression "integrated processing" which is adapted for infrared astronomy and for the infrared detectors signals. The evaluation of this model on the case study "HERSCHEL/PACS" is presented in another technical report

References

[1] G. Held, "data and Image Compression: Tools and Techniques," J. Wiley and Sons, England, 1996

[2] http://www.datacompression.info.

[3] M. Louys, J.L.Starck, S. Mei and F. Murtagh, "Astronomical Image Compression," Astronomy and Astrophysics, page 579-590, May 1999.

[4] M. Datcu and G. Schwarz, "Advanced Image Compression: Specific Topics for Space Applications," DSP'98, and Int. Workshop on DSP techniques for Space Applications 1998.

[5] J. Sanchez and M.P. Canton, "Space Image Processing," CRC Press, 1999.

[6] C. Sterken and M. de Groot, "The Impact of Long-Term Monitoring on Variable Star Research," Proceeding of the NATO Advanced Research Workshop, Belgium, 1993

[7] J. Blommaert et al, "CAM - The ISO Camera," ISO Handbook Volume III, Version 1.2, June 2003.

[8] J.L. Starck, F. Murtagh and A. Bijaoui, "Image Processing and Data Analysis: The Multiscale Approach," Cambridge University Press, USA, 1998.

[9] R.L. Libbey, "Signal and Image Processing Sourcebook," Multiscience Press, USA, 1994.

[10] J. Reichel, "Complexity Related Aspects of Image Compression," PhD, Thesis ,EPFL-No2326, Lausanne, Switzerland, 2001.

[11] S.G. Mallat, "A Theory for multiresolution Signal Decomposition: The Wavelet Representation," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.11,pp.674-693,Jul. 1989.

[12] M. Vetterli, and J. Kovacevic, "Wavelets and Subband Coding," NJ, USA: Prentice Hall, 1995.

[13] W. Sweldens, "The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets," Appl. Comput. Harmon. Anal., vol.3, no.2, pp. 186-200, 1996.

[14] I. Daubechies and W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps," J. Fourier Anal.Appl., vol.4, no.3, pp.245-267, 1998.

[15] H. Zuse, "Software Complexity," NY, USA:Walter de Cruyter, 1991.

[16] S. Winograd, "On the Complexity of Symmetric Filters," in Proc. of IEEE Int. Symp. Cir. Sys., NY, USA, pp.262-265, 1979.

[17] S. Winograd, "Signal Processing and Complexity of Computation," in IEEE ICASSP 80, vol.24, NY, USA, pp. 94-101, 1980.

[18] E. Feig and S. Winograd, "On the Multiplicative Complexity of the Discrete Cosine Transform," in IEEE Trans. on Information Theory, vol. 38, pp. 1387-1397, Jul. 1992.

[19] H.Hakimmashhadi, "Discrete Fourier Transform and FFT," Chap.4, 77-118, NY, USA: Marcel Dekker, 1988.

[20] P. Duhamel and H. H'mida, "New 2^n DCT Algorithms suitable for VLSI Implementation," in IEEE ICASSP 87, NY, USA, pp.1805-1808, 1987.

[21] C.K. Koc, G. Chen and C.K. Chui, "Complexity Analysis of Wavelet Signal Decomposition and Reconstruction," IEEE Trans. on Aerospace and Electronic Systems, vol.30, pp 910-918, Jul.1994.

[22] O.Rioul and P. Duhamel, "Fast Algorithms for Discrete and Continuous Wavelet Transforms," in IEEE Trans. Inf. Theory, vol. 38, Mar. 1992.

[23] K. Froitzheim and H. Wolf, "Knowledge- Based Approach to JPEG Acceleration," in SPIE, the Int. Soc. for Optical Engin., vol 2419, USA pp.318-328, 1995.

[24] K. Lengwehasatit and A. Ortega, "DCT Computation with Minimal Average Number of Operations," in Vis. Com. and IP, VCIP97, USA, Feb. 1997.

[25] K. Lengwehasatit and A. Ortega, "DCT Computation Based on Variable Complexity Fast Approximation," in ICIP'98, USA, pp.95-99, Oct. 1998.

[26] P. Fernandez and A. Ortega, " An Input dependent Algorithm for the Inverse DWT," the 36th Conf. on Sig., Sys and Computers, Asilomar'98, vol.1, pp. 472-476, Nov. 1998.

[27] K. Lengwehasatit and A. Ortega, "Distortion/Decoding Time Tradeoffs in Software DCT-Based Image Coding," in IEEE ICASSP'97, Munich, Germany, 1997.

[28] M.J. Gormish, " Source Coding with Channel, Distortion and Complexity Constraints," PhD Thesis, Stanford University, 1994.

[29] E. Ordentlich, D. Taubman, M.Weinberger, G. Seroussi, and M. Marcellin, "Memory efficient scalable line-based image coding," in IEEE Data Compression Conference, (Snowbird, USA), pp. 218-227, Mar. 1999.

[30] H. Nielsen, "Complexity Evaluation of Signal Processing Algorithms," in EUSIPCO'88, Sig. Proc.IV: Theory and Applications, Netherlands, 1988.

[31] J. Natvig, "Evaluation of Six Medium Bitrate Coders for the Pan-European Digital Mobile Radio System," IEEE Journal on Selected Areas in Communications, vol. 6, pp. 324-331, Feb. 1988.

[32] H. Bischof, A. N. Belbachir, D.C. Hoenigmann, and F. Kerschbaum, "A data reduction concept for FIRST/PACS," in J. B. Breckinridge and P. Jakobsen, editors, UV, Optical, and IR Space Telescopes and Instruments VI. SPIE, Munich, Germany, March 2000.

[33] S. Kullback, "Information Theory and Statistics," Peter Smith, Gloucester, MA, USA, 1978. bibitem [34]Bijaoui A. Bijaoui, "Généralisation de la Transformation d'Anscombe," Technical Report, 1994.

[35] D. Shaw and P. Hodge, "Cosmic Ray Rejection in STIS," Instrument Science Report STIS 98-22, 1998.

[36] W. Wijmans and P. Armbruster,"Data Compression Techniques for Space ApplicationsBijaoui," DASIA'96 , Rome, Italy, May, 1996.

[37] C.E. Shannon,"Technical Journal,", Bell Systems, 1948.

[38] E. Tuncel and K. Rose,"Computation and Analysis of the N-Layer Scalable Rate-Distortion Function,", IEEE Trans. on Information Theory, Vol.49, No.5, May 2003.

[39] J.L. Starck, E.J Candes and D.L. Donoho,"The Curvelet Transform for Image Denoising," IEEE Trans. on Image Processing, vol.11, June 2002.

[40] I.S. Glass, "Handbook of Infrared Astronomy," Cambridge University Press, October 1999.

[41] S. Ott, "Innovative Cosmic Ray Rejection in ISOCAM Data,"

[42] L. Cieplinski, "A Review of Image and Video Coding Standards," Fundamenta Informaticae, vol. 34, pp. 347-367, July 1998.

[43] K. Sayood, "Introduction to Data Compression," Second Edition, Morgan Kaufmann, 2000

[44] N. A. Romero, "Johnson Noise", Junior Physics Laboratory, Massachusetts Institute of Technology Cambridge, Massachusetts 02139, November 1998

[45] . N. Hooge, "1/f Noise is No Surface Effect", Physics Letters, Vol. 29A, Nr. 3, pp. 139-194, April 1969.

[46] . N. Hooge, "1/f Noise Sources", IEEE Transactions on Electron Devices, Vol. 41, Nr.11, pp. 1926-1935, Nov. 1994.

[47] . Grimmett and D. Stirzaker, "Probability and Random Processes," 2nd ed. Oxford, England: Oxford University Press, 1992.

[48] JPEG 2000 Image Coding System" at http://www.jpeg.com/JPEG2000.html, 2000.

[49] C. Chrysafis and A. Ortega, "Line Based, Reduced Memory, Wavelet Image Compression," in IEEE Trans. on Image Processing, 2000.

[50] http://www.iso.vilspa.esa.es

[51] http://www.mpe.mpg.de/projects.html