
This is the **published version** of the bachelor thesis:

Lamas Pérez, José Manuel; Serra-Sagristà, Joan, dir. Quantització avant la lettre. 2023. (958 Enginyeria Informàtica)

This version is available at <https://ddd.uab.cat/record/272810>

under the terms of the  license

Quantització avant la lettre

José Manuel Lamas Pérez

Resumen– La compresión de datos es uno de los aspectos más importantes en un canal de comunicación, y con la ingente cantidad de información que se transmite y almacena hoy en día, se debe seguir investigando y renovando para poder adaptarse a nuevas dificultades. En los últimos años se han presentado diversos estudios en los que demuestran cómo cambiar la arquitectura puede afectar a los resultados obtenidos. En este proyecto se pretende continuar la línea de investigación de estos estudios. Para ello se ha decidido que se intercambie el orden entre la etapa de predicción y cuantización, y los métodos que se aplicarán son JPEG-LS, M-CALIC, CCSDS y cuantización uniforme. Con ello se quiere comprobar cómo estos cambios afectan a los resultados para determinar si se consigue o no una mejora en la eficiencia.

Palabras clave– Compresión de datos, Cuantización uniforme, CCSDS, M-CALIC, JPEG-LS, Cuantización, Predicción.

Abstract– Data compression is one of the most important aspects of a communication channel, and with the enormous amount of information being transmitted and stored today, research and renewal must continue in order to adapt to new challenges. In recent years, several studies have been presented showing how changing the architecture can affect the results obtained. This project aims to continue the research line of these studies. To this end, it has been decided to swap the order between the prediction and quantization stages, and the methods to be applied are JPEG-LS, M-CALIC, CCSDS and uniform quantization. The aim is to test how these changes affect the results to determine whether or not an improvement in efficiency is achieved.

Keywords– Data compression, Uniform Quantization, CCSDS, M-CALIC, JPEG-LS, Quantization, Prediction.



1 INTRODUCTION

In today's world, the amount of data stored and processed in all kind of applications is so big that requires a tool to allow us to facilitate the use of this data. Data compression algorithms provide the necessary techniques to minimize the size of the files, making them easier to transfer and saving resources in terms of time, storage capacity and bandwidth in the network.

Principally, there are two types of data compression, lossless compression and lossy compression [1]. The first type maintains the same data before the compression, but it is rewritten in a more concise way. It is used mainly for text files or software applications, where the files would suffer a malfunction if they were not exactly the original. In the

second type of compression, although the main purpose of data compression is that the data, once decompressed, is as similar as possible to before it was compressed, sometimes, mainly in audio or image files, we can sacrifice part of the data since it contains information imperceptible to the human eye or ear. This way of compression is called irreversible, because once the file is compressed, it is impossible to recover the original file. There is another type, although some researchers include it within the lossy type, called near-lossless compression [2]. In this method, the user selects quantitatively the amount of data to be lost. This can be very useful since it can assure the way in which the parameters of interest can be affected or not affected as the user wants.

Usually, the compression takes place in three stages, decorrelation [3] is the first one and it consists in reducing as much as possible the redundancy or correlation of the data. Shannon [4], described redundancy as “that fraction of a message or datum which is unnecessary and hence repetitive in the sense that if it were missing the message would still be essentially complete, or at least could be com-

- Contact E-mail: JoseManuel.Lamas@autonoma.cat
- Specialization: Information Technology
- Work supervised by: Joan Serra Sagrista (DEIC)
- Academic year: 2022/23

pleted.”. The second one is quantization, which is, basically, a division with the purpose of making the data smaller making it easier to represent. The divisor number is called quantization step, which is important to be balanced, larger numbers will produce redundant values in the output signal and therefore, larger compression capacity, but a lower accuracy in the reconstruction due to the distortion introduced [5]. The third and last stage is entropy coding, which attempts to assign a code-word of the selected code based on the entropy of the data, the concept is to assign the shortest code-words to the symbols that appear most often, and the longest to those that appear less [6].

This document is structure as it follows. In section 2 the objectives and motivation of the project are explained, then, in section 3, is presented the planning and methodology to follow in order to achieve those objectives. Section 4 is a review of the of the currently used techniques, divided in three subsections, according to the three stages of data compression. In sections 5 and 6 are presented the dataset and the software chosen, respectively, along with the reasons to make that choice. Lastly, the section 7 is about the obtained results of the study, followed by the conclusion in section 8.

2 OBJECTIVES

In the first meeting of the CCETT (Centre Commun d'Études de Télécommunications et de Télédiffusion), the three stages of compression were defined like this and remains until this day for historical reasons [7], but even though the acceptance and use of this scheme it presents some difficulties. Usually, the quantization after the decorrelation complicates the reconstruction of the samples at the decompression and it leads to looking for solutions that can make the design less complex when it comes to implementing it.

Recent studies [8], [9] have demonstrated that modifying this architecture can affect to the compression efficiency, achieving in some cases better results than the traditional architecture. This results can suppose a change on the design and implementation of new data compression algorithms, so it is a subject worth of further investigation.

The aim of this project is to study how changing the order of the first (decorrelation) and second (quantization) stages of the compression scheme affects efficiency. In other words, this article seeks to assess whether the added design complexity in the traditional scheme looking for the best efficiency is worth the risk. In figure 1 it is shown the difference between traditional schemes and the proposed scheme for this study. For this purpose it is important to learn about different data compression algorithms and techniques to understand them in order to be able to choose the method that best suits in each situation to obtain the best results.

3 PLANNING AND METHODOLOGY

The chosen methodology for this project is the waterfall method [10]. The main reason for this is because while other methodologies like agile or SCRUM are much more focused on teamwork and require a very good organization

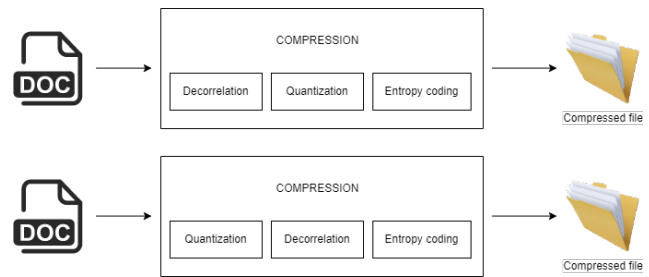


Fig. 1: (Top) Traditional compression scheme. (Bottom) Proposed compression scheme.

so they can be complex to implement, the waterfall methodology is easier in an individual project.

Even though this methodology is well known for software development especially, the working process of this project will be very similar to the model that this method proposes, in which a task will not be started until the previous one was finished. In theory, in this type of methodology each task can only be executed one time, but this project will not follow the waterfall methodology strictly, so if at the end the results are not precisely enough, another technique can be selected.

The work can be defined by setting a few milestones with its respective tasks:

- **Select a software able to perform several techniques of data compression.** To facilitate the process, it was decided to use a software to apply different data compression techniques once the decorrelation and quantization are achieved. For that, it is necessary to do a research and assess the different options.
- **Select the dataset that will be used.** In order to find the dataset that best fits the project, it is important to explore the available datasets, verifying its quality and validity.
- **Recollect information and revise the state of the art.** This part of the work is focused on revising the literature to identify potential techniques that could be of use in our scope.
- **Apply the quantization and decorrelation methods.** Once decided the techniques to use, they have to be implemented in a programming language to subsequently analyze the results offered by each.
- **Analyze the results and make conclusions.** In this stage of the project, the performance of the implemented techniques is evaluated to extract valuable conclusions on the work performed.
- **Document the process.** The last phase of the process is oriented to the documentation of all the tasks carried out during this work.

The first two tasks will require a little research to achieve the objective, so they can be done in one week. Gathering the information for the third one can be more difficult, due to the need of revising similar projects to assure the techniques used are the most appropriate, two weeks or three are needed. The fourth and fifth task should be the large part of

the project since they require code implementation, an estimated time of six weeks is suitable. Finally, for the sixth task two weeks will be correct. These times are all estimations based on previous works, giving a five week margin to manage unexpected events taking into account the first week of February as the end date of the project. In figure 2 it is shown graphically the planning of the project.

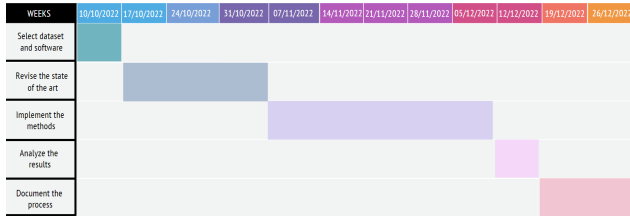


Fig. 2: Gantt diagram of the project.

4 STATE OF THE ART

The need for reducing the size of the files has resulted in a huge catalogue of different ways to complete the compression in an attempt to find the best for each type of file. As it is said before, this paper is going to focus on techniques for decorrelation and quantization. Despite that, a section of data compression techniques is included since according which one is used, it can affect the results, so the ones that fit best for the analysis have to be selected.

4.1 Decorrelation

Decorrelation can be accomplished by many techniques, but the most applied are linear prediction and transformation. Linear prediction can be completely reversible, the way that it works is that for each sample, a prediction of its value is formed from a weighted sum of neighboring samples [11]. The prediction residual, has much lower entropy than the original data. In the other hand, transformation techniques are frequently employed on lossy compression systems. The most popular method is a variant of the discrete Fourier transformation (DFT) called discrete wavelet transformation (DWT). One of the main reasons for that is that the DWT allow to recover the image with very few coefficients once these are quantized [12].

4.2 Quantization

The form in which the data is presented affects the way the quantization is done. If the symbols of the input data appear with the same probability the quantization should be uniform, it means that the quantization step has to be the same for all the symbols in the data. This situation generally takes place when the source of the data is memory-less. In the other hand, if the symbols of the data are not uniform distributed, the quantization step should be adjusted for each symbol. For those that appear more frequently, the quantization step is small, while for those that appear less, it should be larger.

4.3 Data compression techniques

When it comes to data compression techniques, the standard proposed from the Joint Photographic Experts Group (JPEG) has been widely accepted for still image compression. The main goal for JPEG was to develop an international standard for compression and decompression of continuous-tone, still frame and color images. It has four modes: lossless, sequential, progressive and hierarchical. The first one is a lossless mode and the other three are lossy modes [13]. With the time, a new standard of JPEG was created, called JPEG-LS [14] focused on provide efficient lossless compression at a reasonable complexity.

In July 1995, at the ISO/SC29/WG1 meeting, the Context-based, Adaptive Lossless Image Coding (CALIC) was proposed to ISO/JPEG with the intention to be the next international standard for lossless compression, having the lowest bit rates in six of seven image classes: medical, aerial, pre-press, scanned, video and compound document, and the third lowest bit rate in computer generated images [15]. Years later, Enrico Magli, proposed a CALIC based technique for hyperspectral images, calling it M-CALIC [16], outperforming some of the techniques used at the moment. M-CALIC has both low encoder complexity and high performance, and it does not provide any error resilience.

Another important technique, specially for multispectral and hyperspectral images, is the one published by the Consultative Committee for Space Data Systems (CCSDS). These images are so important because they are the output of the sensors used for geoscience and Earth observation [17] whose use has significantly grown in the last years. Currently, the latest standard supporting this type of images is the CCSDS 123.0.B-2 [18], which implements lossless compression and near-lossless compression.

5 DATASET

Regarding the datasets, there are a few well known examples such as the calgary corpus, replaced later with the canterbury corpus used specially for data compression. Of course, the selected dataset should be chosen taking into account the objective of the study. As this study is more centered on image compression techniques, the canterbury corpus is not an option since it was created for text compression tests. Also datasets created for image compression tests exist, like the IRMA corpus or the silesia corpus, focused on medical photos, but due to the techniques chosen in this study the selected dataset is the CCSDS test data corpus [20], used for developing the previous standard of the CCSDS-123.0-B-2.

This corpus contains hyperspectral and multispectral images, which as said before, are the main objective of the CCSDS standard, and as shown in [19], JPEG-2000 can also perform well with this type of photos. Hyperspectral and multispectral images are formed by multiple spectral bands, instead of the traditional RGB system. The number of bands depends on the sensor and it can be up to a few thousands, where the pixels have a different value in each band. Thanks to that, an image can contain a lot of information with great detail, which make the hyperspectral images very useful in the spectrometry of the vegetation and minerals of the Earth [21].

TABLE 1: SELECTED DATASETS FOR THE STUDY

Instrument	Images
AVIRIS	Yellowstone Scene: 0, 3, 10, 11, 18 (Calibrated) 244 bands, 512 rows, 677 columns
	0, 3, 10, 11, 18 (Non calibrated) 244 bands, 512 rows, 680 columns
M3	Target A Target B Target C 260 bands, 512 rows, 640 columns
AIRS	gran16 gran120 gran129 gran193 1501 bands, 135 rows, 90 columns
CRISM	181
	182
	183
	545 bands, 420 rows, 320 columns

The corpus includes images from different instruments, in table 1 the images selected for this experiment are presented. In the table are shown ten images from AVIRIS (Airbone Visible/Infrared Imaging Spectrometer) instrument, five of these images are calibrated Yellowstone scenes with 244 bands, 512 rows and 677 columns and the other five uncalibrated Yellowstone scenes with 244 bands, 512 rows and 680 columns, three images from M3 instrument, with 260 bands, 512 rows and 640 columns, four images from AIRS (Atmospheric Infrared Sounder), with 1501 bands, 135 rows and 90 columns and three images from CRISM (Compact Reconnaissance Imaging Spectrometer for Mars) with 545 bands, 420 rows and 320 columns. All the images are in 16 bits per pixel per component (bpppc) and are in .raw format

6 SOFTWARE

In order to perform the data compression techniques mentioned before, a tool able to execute them is needed. The techniques can be implemented too, but as the analysis of the algorithms is not the objective of this paper, it is considered that it is better to invest the time in the main goal of the project. There are several softwares or libraries that can perform these techniques, but they are usually specialized in only one technique.

For that reason, the software that will be used in this study is the Experiment Notebook (ENB) [22]. As its page says, the ENB is an open-source library focused on helping obtaining and reporting computer-based experimental data, allowing the user to focus only in the experiment he aims to perform without having to be concerned about other details.

It consists in an easy Python interface through the Linux terminal and it can be personalized by adding or changing the code scripts. Among its features, it allows to perform image compression experiments very easily. Once the user has processed the images, which in this case involves quantization and prediction, the user only has to choose a method among those the ENB can perform. The results can be pre-

sented in different formats, like plots in PDF or PNG or tables in .tex. Besides, it not only compresses the image, but also can show more details like the compression and decompression time, the compression ratio or the bits per pixel per component (bpppc), according on what the user chooses.

7 EXPERIMENTS

To begin with, some tests were run to get acquainted with the ENB. The methods with which these tests were run are described in section 4.3, the JPEG-LS, the CCSDS and the M-CALIC. In order to run the software with these techniques, a plugin called lossless-compression must be installed. That plugin includes a script in python to initialize and configure the ENB execution, such as the codecs with its respective parameters and the folder containing the data.

The way in which the quantization is introduced for the presented techniques in the traditional scheme is through the peak absolute error (PAE). The PAE is a parameter that measures the quality of the compression using the original image and the compressed image. In these methods the user can select the desired PAE, which sets a bound for the method forcing it to have a maximum admissible error, to determine which are the quantization steps.

As said before, the objective of this study is to move the quantization stage before the decorrelation stage in the compression scheme, so the first step needs to be the quantization. As a first approach, an initial script in python to perform uniform quantization was implemented, with the idea of understanding the concept. The script was very simple; the image was read and then passed through a *for* loop to divide the value of the pixels several times by different quantization steps.

This script was run with some images outside the dataset and the results were successful, but when it was tried with the images from the CCSDS corpus it did not work due to the impossibility of saving .raw images. However, this was not a problem, since ENB includes a module for uniform quantization, in which the user can select the quantization step and execute it.

With these first simulations completed, the ENB was used so the .raw images could be handled. In the case of the simulations with the traditional scheme, the quantization steps were selected in the proper ENB script, since all the techniques selected count with near-lossless implementation and a predictor in the decorrelation stage. In the other hand, for the scheme proposed in this study, the images were first passed through the ENB uniform quantization module mentioned before, specifying the quantization step desired. In order to decide the range values of this parameter, a research was done to see what are the values that are used in the practice, so this work will follow the example of article [17], from 1 (lossless) to 32. Another parameter the user can select in the ENB is the number of repetitions of each execution to obtain a more reliable value of the compression time. After a few tests of this parameter, it was verified that with the number of repetitions set in 10, the results were reliable enough.

In figures 3, 4, 5, 6 and 7 the results for the before mentioned techniques with different quantization steps are presented, tables 2 and 3 and 4, 5 and 6 complement these

TABLE 2: VALUES FOR AVIRIS CALIBRATED

TRADITIONAL SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	6,38	4,08	3,34	2,64	1,99	1,41	17,15	15,24	15,17	13,70	11,82	10,11
CCSDS	4,16	2,09	1,48	1,01	0,68	0,44	10,15	10,90	11,48	10,91	9,88	9,33
M-CALIC	3,86	1,85	1,30	0,86	0,54	0,34	62,23	59,34	58,55	57,62	57,07	56,42
PROPOSED SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	6,38	5,42	4,47	3,61	2,83	2,13	17,15	16,98	15,38	13,74	12,52	10,76
CCSDS	4,16	3,24	2,40	1,74	1,29	0,99	10,15	11,68	11,36	10,13	10,50	10,08
M-CALIC	3,86	2,98	2,24	1,77	1,46	1,24	62,23	60,86	61,60	59,71	57,98	56,70

TABLE 3: VALUES FOR AVIRIS UNCALIBRATED

TRADITIONAL SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	8,61	6,30	5,46	4,59	3,67	2,90	19,11	19,19	18,00	16,69	14,76	13,58
CCSDS	6,68	4,38	3,60	2,79	1,99	1,36	11,88	11,04	11,14	11,37	10,70	10,44
M-CALIC	6,04	3,72	2,92	2,11	1,39	0,87	65,00	61,67	60,89	59,80	58,73	58,05
PROPOSED SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	8,61	7,60	6,63	5,68	4,74	3,85	19,11	18,36	18,43	17,92	17,93	13,95
CCSDS	6,68	5,55	4,56	3,63	2,77	2,04	11,88	12,08	11,19	11,84	11,66	10,50
M-CALIC	6,04	5,03	4,06	3,17	2,39	1,85	65,00	63,62	62,60	61,74	60,43	59,23

TABLE 4: VALUES FOR M3

TRADITIONAL SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	3,66	1,78	1,46	1,28	1,21	0,96	12,70	11,22	11,77	10,55	9,40	6,03
CCSDS	6,49	4,17	3,74	3,05	2,17	1,41	7,45	8,45	8,02	7,77	7,89	7,39
M-CALIC	5,12	2,82	2,10	1,47	0,93	0,49	65,04	62,10	61,22	60,12	60,17	58,09
PROPOSED SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	3,66	2,75	1,99	1,52	1,30	1,21	12,70	11,81	10,08	9,73	8,63	7,95
CCSDS	6,49	5,49	4,52	3,71	2,93	1,43	7,45	7,42	7,37	7,33	7,21	7,18
M-CALIC	5,12	4,11	3,32	2,68	2,06	1,20	65,04	63,42	62,21	61,17	60,06	57,82

TABLE 5: VALUES FOR AIRS

TRADITIONAL SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	6,19	3,91	3,14	2,40	1,72	1,12	4,64	4,33	4,15	3,62	2,70	2,25
CCSDS	4,41	2,21	1,57	0,99	0,57	0,36	2,64	2,47	2,52	2,76	2,39	2,45
M-CALIC	4,23	2,06	1,43	0,86	0,46	0,27	14,66	14,13	13,88	13,78	13,55	13,41
PROPOSED SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	6,19	5,22	4,27	3,36	2,54	1,82	4,64	4,30	4,18	3,94	3,31	2,85
CCSDS	4,41	3,42	2,52	1,80	1,31	0,95	2,64	2,76	2,68	2,59	2,62	2,65
M-CALIC	4,23	3,31	2,53	1,99	1,59	1,27	14,66	14,54	14,26	14,08	13,82	13,50

TABLE 6: VALUES FOR CRISM

TRADITIONAL SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	5,56	3,35	2,64	2,04	1,61	1,29	19,08	18,34	17,64	16,63	15,23	12,96
CCSDS	6,95	4,69	3,97	3,26	2,41	1,69	13,25	12,62	12,62	12,38	12,28	12,30
M-CALIC	7,42	4,98	4,12	3,27	2,45	1,70	66,38	65,28	64,17	59,08	58,25	57,47
PROPOSED SCHEME												
TECHNIQUE	BPPPC						SECONDS					
	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32	QS 1	QS 2	QS 4	QS 8	QS 16	QS 32
JPEG-LS	5,56	4,62	3,68	2,84	2,14	1,65	19,08	18,68	17,62	16,87	15,48	14,02
CCSDS	6,95	5,96	4,99	4,08	3,34	1,97	13,25	13,13	12,84	12,68	12,51	11,96
M-CALIC	7,42	6,33	5,26	4,25	3,37	2,57	66,38	63,90	62,30	60,79	59,53	58,14

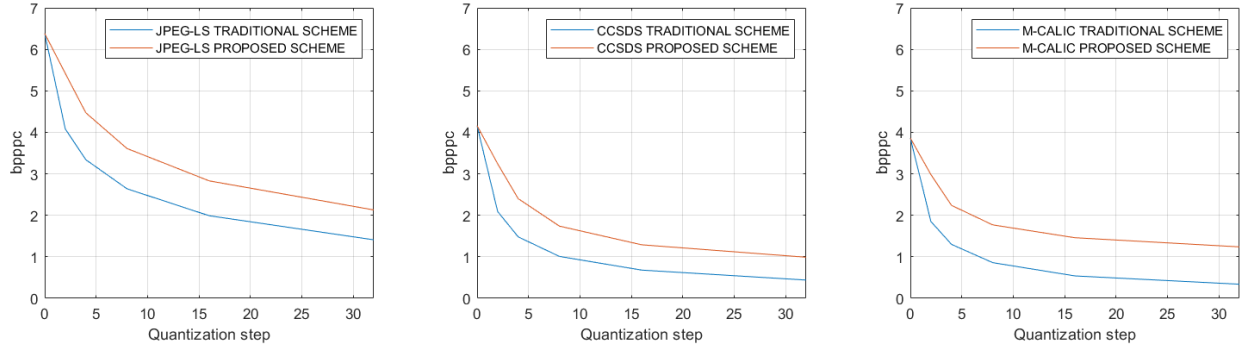


Fig. 3: Comparison between the traditional and the proposed scheme for the selected methods with calibrated images from AVIRIS.

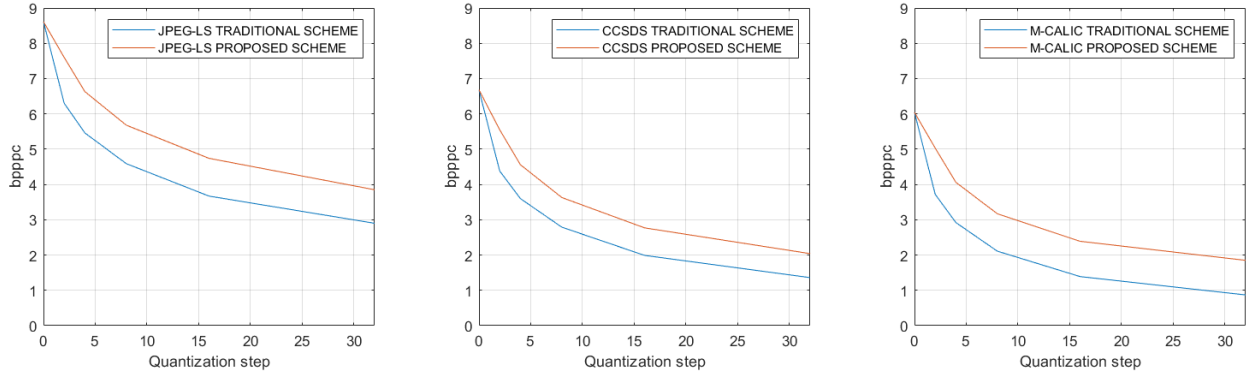


Fig. 4: Comparison between the traditional and the proposed scheme for the selected methods with uncalibrated images from AVIRIS.

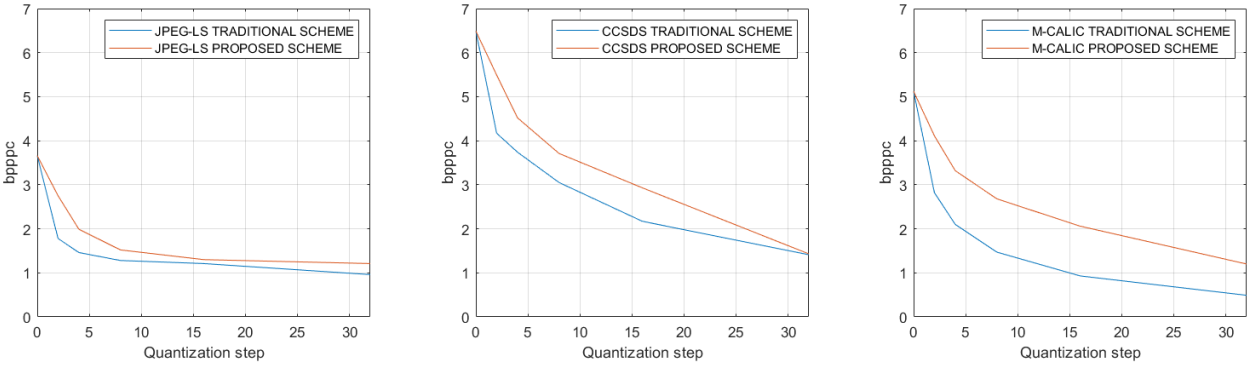


Fig. 5: Comparison between the traditional and the proposed scheme for the selected methods with images from M3.

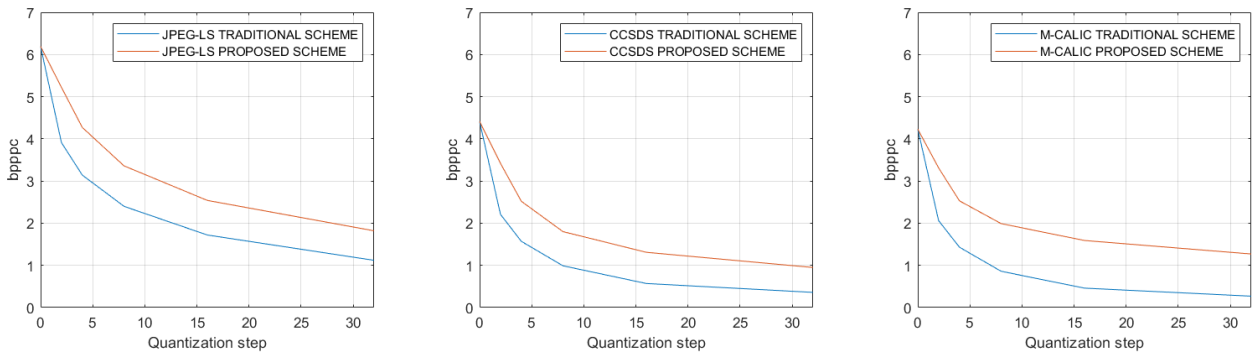


Fig. 6: Comparison between the traditional and the proposed scheme for the selected methods with images from AIRS.

figures respectively. The plotted values are the mean of the compressed data rate measured in bits per pixel per component of the images from the different datasets. As expected, the larger the quantization step, the lower is the compressed

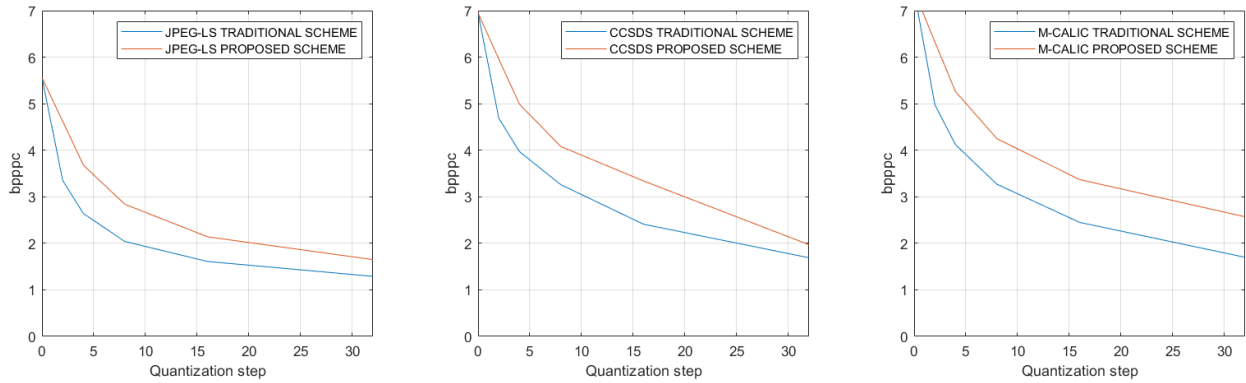


Fig. 7: Comparison between the traditional and the proposed scheme for the selected methods with images from CRISM.

data rate, as said before, in the practice it is important to find a trade-off between compressed data rate and the value of the quantization step in order to not get redundancy in the data. The optimal quantization step depends on the the purpose of each case.

Another interesting metric when it comes to compare the efficiency of the methods is the compression time. In tables 2, 3, 4, 5 and 6 it is shown the average compression times measured in seconds along with the compressed data rate in bpppc for both schemes with the different quantization steps.

Looking at the presented results above, analyzing the compressed data rate, can be said that in general, the M-CALIC has the best performance of the three. It is the technique with the lowest compressed data rate for all the datasets, except for the M3 and the CRISM, being the technique with the highest compressed data rate in this last dataset. However, the CCSDS achieves very similar results in almost every datasets, and depending on the quantization step used, with the proposed architecture it is possible to obtain a better result than the M-CALIC in terms of the bpppc, like for AVIRIS calibrated, AIRS and CRISM. The difference between JPEG-LS with the CCSDS, which is the second technique with the best performance, is bigger than the difference between the CCSDS with the M-CALIC, having the JPEG-LS the highest compressed data rate in almost every dataset, but for the M3 and the CRISM is the one with the lowest bpppc. The results between datasets are very similar, the AVIRIS uncalibrated results are slightly high than the others but despite that, the difference is not very big.

Observing the time results, the CCSDS is clearly the one that performs better in this aspect, followed by the JPEG-LS and then the M-CALIC. It can be observed that the quantization also affects the compression time, being faster the more the quantization step increases. The more affected ones by the quantization are the JPEG-LS and the M-CALIC, reducing around five seconds or more from lossless compression to the last quantization step, allowing cases where the JPEG-LS gets to be as fast or even faster than the CCSDS. It is worth noticing that the CCSDS is barely affected, reducing only one second in the cases that is most affected and even are some other cases in which is not affected at all. It can be seen that the average times for each technique are around ten seconds for the CCSDS, between twenty and ten with the quantization for the JPEG-LS and

between sixty five and fifty seven with the quantization for the M-CALIC, except for the AIRS dataset, which achieve the fastest times of the study.

Comparing both schemes, which is the object of the study, looking at the graphics can be seen that the traditional architecture performs slightly better than the proposed one for all the datasets. However, the difference between the compressed data rate from one scheme and the other is not always the same. In the case of the M3 dataset, specially for the JPEG-LS and the CCSDS there are points where the results with the proposed architecture are as good as with the traditional one, but for the M-CALIC the difference is one of the biggest of the study. In general, the biggest difference is from the M-CALIC, being the more distant technique to achieve a similar performance in both schemes. JPEG-LS has points where the results from both schemes coincide, such as the commented before and in the CRISM dataset for larger quantization steps. The CCSDS is the one with the lowest difference between the two schemes, as in the JPEG-LS, in the M3 and in the CRISM datasets the CCSDS have points where the performance is approximately the same in the two architectures, specially for larger quantization steps. In terms of time, it can not be said that one scheme performs better than the other. In all the datasets the compression times are almost the same, with less than a second of difference between the two architectures for the same quantization step in the majority of the cases.

8 CONCLUSIONS

Analyzing the results obtained it seems that the effect of moving the quantization before the prediction depends on the dataset and quantization step used. As it is shown in the different tables and graphics presented in the study, in the majority of the cases the traditional architecture performs better than the proposed scheme, but there are some other cases where the resulting compressed data rate of moving the quantization stage can be as good as in the traditional architecture. Regarding the time results, there is no significant difference between both scenarios. Taking this into account, it can be said that the traditional architecture achieve better results than the proposed one.

However, the results presented in this study show that different factors, such as the dataset and techniques selected, can substantially affect the compression efficiency, obtaining very different results for different datasets using the

same techniques and vice versa. As a result of that, a possible future work to continue with this study could be to keep testing with different techniques, both in the quantization stage and in the prediction stage or even to expand the dataset, in order to match the results of the commented studies.

ACKNOWLEDGMENTS

I would like to thank my tutor, Joan Serra Sagrista, for the recommendations, help and support during the developing of this project.

I also would like to thank my friends and family for the love and moral support they gave me these last years.

Finally, I would like to thank Miguel Hernández-Cabronero, creator of the ENB, for the help with the software when technical problems appeared.

REFERENCES

- [1] "¿Qué es la compresión de datos? (Definición, importancia y métodos)" , historiadelaprensa.com. [Online]. Available: <https://historiadelaprensa.com/compresion-de-datos>.
- [2] Rashid Ansari, Nasir Memon and Ersan Ceran, "Near-lossless image compression techniques"
- [3] J. Wang, M. Zhang and S. Tang, "Spectral and spatial decorrelation of Landsat-TM data for lossless compression," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 5, pp. 1277-1285, Sept. 1995, doi: 10.1109/36.469492.
- [4] C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," Urbana, IL: Illinois Press, 1949
- [5] O. Kadri and Z. -E. Baarir, "Still image compression using Curvelets and logarithmic scalar quantization technique," 2016 5th International Conference on Multimedia Computing and Systems (ICMCS), 2016, pp. 645-649, doi: 10.1109/ICMCS.2016.7905667.
- [6] S. M. S. Jalaaliddine, C. G. Hutchens, R. D. Stratton and W. A. Coberly, "ECG data compression techniques-a unified approach," in *IEEE Transactions on Biomedical Engineering*, vol. 37, no. 4, pp. 329-343, April 1990, doi: 10.1109/10.52340.
- [7] Graham Hudson, Alain Léger, Birger Niss, István Sebestyén, Jørgen Vaaben, "JPEG-1 standard 25 years: past, present, and future reasons for a success," *J. Electron. Imag.* 27(4) 040901 (31 August 2018) <https://doi.org/10.1117/1.JEI.27.4.040901>
- [8] Chatziantoniou, P., Tsigkanos, A., Theodoropoulos, D., Kranitis, N., Paschalis, A. (2022). An Efficient Architecture and High-Throughput Implementation of CCSDS-123.0-B-2 Hybrid Entropy Coder Targeting Space-Grade SRAM FPGA Technology. *IEEE Transactions on Aerospace and Electronic Systems*, 58(6), 5470-5482.
- [9] Sánchez, A., Blanes, I., Barrios, Y., Hernández-Cabronero, M., Bartrina-Rapesta, J., Serra-Sagristà, J., Sarmiento, R. (2022). Reducing Data Dependencies in the Feedback Loop of the CCSDS 123.0-B-2 Predictor. *IEEE Geoscience and Remote Sensing Letters*, 19, 1-5.
- [10] "El modelo en cascada: desarrollo secuencial de software", ionos.es. [Online]. Available: <https://www.ionos.es/digitalguide/paginas-web/desarrollo-web/el-modelo-en-cascada/>
- [11] Ming Yang and N. Bourbakis, "An overview of lossless digital image compression techniques," 48th Midwest Symposium on Circuits and Systems, 2005., 2005, pp. 1099-1102 Vol. 2, doi: 10.1109/MWSCAS.2005.1594297
- [12] D. Salomon, "Data Compression: The Complete Reference," Northridge, CA: California University, 2006
- [13] A. Subramanya, "Image compression technique," in *IEEE Potentials*, vol. 20, no. 1, pp. 19-23, Feb-March 2001, doi: 10.1109/45.913206.
- [14] M. W. Marcellin, M. J. Gormish, A. Bilgin and M. P. Boliek, "An overview of JPEG-2000," *Proceedings DCC 2000. Data Compression Conference*, 2000, pp. 523-541, doi: 10.1109/DCC.2000.838192.
- [15] Xiaolin Wu and N. Memon, "CALIC-a context based adaptive lossless image codec," 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings, 1996, pp. 1890-1893 vol. 4, doi: 10.1109/ICASSP.1996.544819.
- [16] E. Magli, G. Olmo and E. Quacchio, "Optimized on-board lossless and near-lossless compression of hyperspectral data using CALIC," in *IEEE Geoscience and Remote Sensing Letters*, vol. 1, no. 1, pp. 21-25, Jan. 2004, doi: 10.1109/LGRS.2003.822312.
- [17] M. Hernández-Cabronero et al., "The CCSDS 123.0-B-2 "Low-Complexity Lossless and Near-Lossless Multispectral and Hyperspectral Image Compression" Standard: A comprehensive review," in *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 4, pp. 102-119, Dec. 2021, doi: 10.1109/MGRS.2020.3048443.
- [18] Kiely, Aaron B., et al. "The new CCSDS standard for low-complexity lossless and near-lossless multispectral and hyperspectral image compression." *International Workshop on On-Board Payload Data Compression*. No. GSFC-E-DAA-TN61202. 2018.
- [19] Q. Du and J. E. Fowler, "Hyperspectral Image Compression Using JPEG2000 and Principal Component Analysis," in *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 2, pp. 201-205, April 2007, doi: 10.1109/LGRS.2006.888109.
- [20] -, CCSDS TEST DATA, [Online] Available: <https://cwe.ccsds.org/sls/docs/Forms/AllItems.aspx?RootFolder=%2Fsls%2Fdocs%2FSLS%2DDDC%2F123%2E0%2DB%2DInfo%2FTestData>

- [21] Avid Roman-Gonzalez, Natalia Indira Vargas-Cuentas. Análisis de imágenes hiperespectrales. Revista Ingeniería y Desarrollo, 2013, Año 9 (Nº 35), pp.14-17. fahal-00935014f
- [22] M. Hernández-Cabronero, "Experiment Notebook user manual", 2020 [Online] Available: <https://miguelinux314.github.io/experiment-notebook/index.html>