Publications of the Astronomical Society of the Pacific, 136:114506 (11pp), 2024 November https://doi.org/10.1088/1538-3873/ad8b69

Lossy Compression of Integer Astronomical Images Preserving Photometric Properties*

Òscar Maireles-González, Joan Bartrina-Rapesta, Miguel Hernández-Cabronero, and Joan Serra-Sagristà
Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Bellaterra, 08193, Spain; oscar.maireles@uab.cat
Received 2024 July 22; accepted 2024 October 25; published 2024 November 14

Abstract

Observatories are producing astronomical image data at quickly increasing rates. As a result, the efficiency of the compression methods employed is critical to meet the storage and distribution requirements of both observatories and scientists. This paper presents a novel lossy compression technique that is able to preserve the results of photometry analysis with high fidelity while improving upon the state of the art in terms of compression performance. The proposed compression pipeline combines a flexible bi-region quantization scheme with the lossless, dictionary-based, LPAQ9M encoder. The quantization process allows compression performance and photometric fidelity to be precisely tailored to different scientific requirements. A representative data set of 16-bit integer astronomical images produced by telescopes from all around the world has been employed to empirically assess its compression-fidelity trade-offs, and compare them to those of the de facto standard Fpack compressor. In these experiments, the widespread SExtractor software is employed as the ground truth for photometric analysis. Results indicate that after lossy compression with our proposed method, the decompressed data allows consistent detection of over 99% of all astronomical objects for all tested telescopes, maintaining the highest photometric fidelity (as compared to state of the art lossy techniques). When compared to the best configuration of Fpack (Hcompress lossy using 1 quantization parameter) at similar compression rates, our proposed method provides better photometry precision: 7.15% more objects are detected with magnitude errors below 0.01, and 9.13% more objects with magnitudes below SExtractor's estimated measurement error. Compared to the best lossless compression results, the proposed pipeline allows us to reduce the compressed data set volume by up to 38.75% and 27.94% while maintaining 90% and 95%, respectively, of the detected objects with magnitude differences lower than 0.01 mag; and up to 18.93% while maintaining 90% of the detected objects with magnitude differences lower than the photometric measure error.

Unified Astronomy Thesaurus concepts: Astronomy data reduction (1861); Astronomy data analysis (1858); Astronomy software (1855)

1. Introduction

Photometry (Henden & Kaitchuck 1982; Romanishin 2006; Budding & Demircan 2007; Evans 2010) is the measurement of electromagnetic radiation, in the form of photons, emitted by astronomical objects. Through the photoelectric effect, the photons received by the telescope detector, i.e., the CCD, are converted into an electrical signal for each pixel of the CCD, resulting in the number of counts for each pixel in the image. The number of counts can be converted into a term called magnitude, which quantifies the brightness of the object. The first to classify stars by magnitude was Hipparchus in 130 BC, with magnitude 1 stars being the brightest observed in the night sky and magnitude 6 stars being the faintest he could distinguish. This scale is maintained to this day and was refined by Pogson in 1856, defining a magnitude 1 star as 100 times brighter than a magnitude 6 star:

$$m_1 - m_2 = 2.5 \log_{10} \frac{F_2}{F_1} \quad (1)$$

where m_i is the magnitudes of the stars and F_i is the received photon fluxes. If $m_1 = 1$ and $m_2 = 6$, F_1 is 100 times larger than F_2. By measuring the magnitude of an astronomical object for different wavelengths, information about the size, distance, and temperature of the star is obtained.

Observatories around the world are generating image data at an ever-increasing pace (Kremer et al. 2017; Grange et al. 2022;

* This work was supported in part by the Spanish Ministry of Science and Innovation (MICINN) and by the European Regional Development Fund (FEDER), funded by MCIN/AEI/10.13039/501100011033/FEDER, UE, under grant PID2021-125258OB-I00; also by the Catalan Government under grant SGR2021-00643; and also by the "PhotSat" project under the Institute for Space Studies of Catalonia (IEEC) and the Catalan Government.

Poudel et al. 2022). Gathered data needs to be stored for years and even decades to sustain current and future scientific studies, given that nowadays more publications are based on archive data than on newly acquired data (S. T. S. Institute 2024). To control and reduce storage and transmission costs, efficient data compression methods need to be applied. Lossless compression offers perfect photometric data integrity, at the cost of limited compression performance (Maireles-González et al. 2023). When data needs to be compressed further to meet more stringent requirements or reduce overall cost, lossless compression is not sufficient, and lossy compression is required (Kremer et al. 2017; Grange et al. 2022; Poudel et al. (2022). In this case, it is paramount to select compression configurations that maintain sufficient image quality when photometry is the main focus of the observatory. Multiple authors have considered errors below 0.01 mag to be small enough for several scientific applications such as searching for transits of extra-solar planets, asteroseismology, Sun-like stars, and variability of faint galactic and extra-galactic objects over large fields of view (Borucki & Lasher 2001; López-Morales 2006; Blake et al. 2008; McGraw et al. 2009; Bono et al. 2010; Di Cecco et al. 2010; Loh et al. 2012; Crawford 2017; Ivănescu et al. 2021).

In 2011, NASA led the development of Fpack Pence et al. (2021), a suite of lossless and lossy compression tools for astronomical images. Fpack has been deployed in a large majority of observatories (European Southern Observatory 2024; Las Cumbres Observatory 2024; Isaac Newton Group Archive 2024), typically in combination with the Flexible Image Transport System (FITS; Hanisch et al. 2001) data format. Fpack supports several compression modes: Rice (Rice et al. 1993), Hcompress (White et al. 1992), Gzip (Gailly & Adler 1992), and PLIO (IRAF Tody 1993). Of these, Hcompress tends to produce the best lossless compression performance, while Rice achieves comparable results with three times lower computational complexity (Maireles-González et al. 2023). Therefore, these two modes are hereafter considered representative of this de facto standard for performance comparisons. Although Fpack is able to apply lossy compression, parameters controlling the introduced errors are not directly related to the photometric analysis process and have no explicit scientific meaning. The introduced loss does not distinguish between background and foreground values, nor does it adjust to object profiles. This is also true about lossy compression methods not designed for astronomical imagery.

The main contribution of this paper is a novel lossy compression method designed to improve upon the state of the art compression performance while being able to adapt to the photometry fidelity requirements of any observatory and telescope. The proposed pipeline consists of a flexible, bi-region quantization scheme followed by an adaptive entropy coder. The quantization scheme takes a threshold t parameter that determines the classification of each pixel as background or foreground. These regions are quantized using independent quantization steps, respectively, controlled with the q_b and q_f parameters. The resulting quantization indices and side information are compressed using the LPAQ9M entropy coder (Mahoney 2024), a dictionary-based technique. By adjusting these parameters, compression performance can be traded for photometric fidelity.

Exhaustive experiments have been performed to support and validate the proposed pipeline. In these experiments, representative corpora of astronomical images from different telescopes are considered, and a recent version of the SExtractor (Bertin & Arnouts 1996) software is used as the ground truth for all photometric results. In these experiments, the selection of optimal t, q_b, and q_f parameter values is discussed, and their performance is compared to the best configurations of Fpack.

Section 2 provides a concise review of work related to the area of lossy compression of astronomical image data. Section 3 introduces our novel lossy compression pipeline. The experiments performed to evaluate and compare this pipeline are provided in Section 4. Finally, Section 5 adds some concluding remarks and describes future work.

2. Related Work

Lossy compression methods specific for astronomical image data have been described in the literature (Press 1992; Shamir & Nemiroff 2005; Pence et al. 2010; Pulido et al. 2020). In most cases, information loss is introduced in a quantization stage, whose coarseness determines the final compressed data bit rate and the accuracy of successive photometric analysis.

Several of the most popular lossy compression algorithms for astronomical images have been developed to comply with the FITS data format, which can store simultaneously images and tables in a single file with multiple extensions, along with header information for each one (Hanisch et al. 2001). FITSPRESS (Press 1992) was an early, if not the earliest, lossy compressor supported in this format. Based on a user-defined loss level, the locations of pixels above a threshold are compressed using run-length encoding and Huffman Coding. The remaining pixels undergo a two-dimensional wavelet transform, followed by a uniform scalar quantization stage and a run-length encoder. The PHOTZIP (Shamir & Nemiroff 2005) codec was designed to reduce the impact of compression on photometry analysis by losing only background information. First, pixel values above a threshold are preserved without loss, while a uniform quantizer is applied to all other pixels. This threshold is computed based on the value of a window of neighboring pixels and a user-defined parameter. Finally, entropy coding is performed using the well-known gzip or bzip2 codecs.

More recently, a pipeline similar to Press (1992) based on spatial CDF wavelets and quantization has also been proposed by Pulido et al. (2020). Alternatively to Press (1992),

Pulido et al. (2020) uses a highly efficient floating-point-to-integer quantization method that shifts the decimal of the floating-point representation to the right, to preserve the integer portion and truncate the fractional half. This method is available in HDF5 Folk et al. (2011).

Fpack, also compliant with the FITS data format, is used by virtually many major observatories such as the Isaac Newton Group Archive (2024), Las Cumbres Observatory (2024) or the European Southern Observatory (2024). The Hcompress (White et al. 1992) and Rice modes (Pence et al. 2010) are the de facto standard for compressing astronomical image data. Lossy compression can be performed by applying an integer-to-float conversion stage in the encoder. Float-to-integer conversion is automatically applied by the software when using the Rice technique on floating-point data. The resulting image samples after applying these conversions are defined as

$$I_{x,y} = \text{round}\left(\frac{I'_{x,y} - \delta}{S(q)}\right). \tag{2}$$

Here, $I'_{x,y}$ is the floating-point value, $S(q)$ is the quantization step size that controls the distortion introduced, and δ is an optional zero-point offset. A user-defined parameter q is used to control the quantization step size $S(q)$ either directly or relative to the rms noise σ, where

$$S(x) = \begin{cases} \sigma/x & \text{if } x > 0 \\ |x| & \text{if } x < 0. \\ 1 & \text{if } x = 0 \end{cases} \tag{3}$$

When the Hcompress mode is enabled in Fpack, the 2×2 H-transform is applied before entropy coding as follows:

$$\begin{aligned} h_{x,y} &= \frac{I_{x+1,y+1} + I_{x+1,y} + I_{x,y+1} + I_{x,y}}{S(s)} \\ h_{x+1,y} &= \frac{I_{x+1,y+1} + I_{x+1,y} - I_{x,y+1} - I_{x,y}}{S(s)} \\ h_{x,y-1} &= \frac{I_{x+1,y+1} - I_{x+1,y} + I_{x,y+1} - I_{x,y}}{S(s)} \\ h_{x+1,y-1} &= \frac{I_{x+1,y+1} - I_{x+1,y} - I_{x,y+1} + I_{x,y}}{S(s)} \end{aligned}, \tag{4}$$

where s is a user-defined parameter that determines the quantization step size, and S is defined in Equation (3).

Note that the lossy processes described in Equation (2) and Equation (4) can be applied sequentially. In this case, the user must choose the value of both q and s.

3. Proposed Technique

This paper proposes a novel lossy compression pipeline consisting of a flexible bi-region quantization stage followed by the LPAQ9M algorithm of the PAQ family. A diagram of the proposed compression pipeline is presented in Figure 1 and its main stages are described next in Sections 3.1 and 3.2.

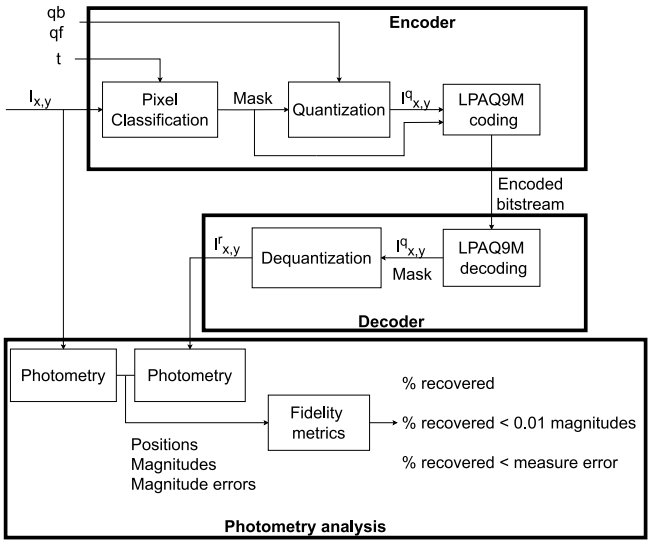

Figure 1. Proposed compression pipeline comprising coding and decoding stages. Posterior photometry and scientific analysis stage is also included.

3.1. Bi-region Quantizer

The proposed quantization stage employs three user-defined parameters: the threshold t, the foreground quantization step size q_f, and the background quantization step size q_b. This pixel classification follows a strategy similar to Press (1992) and Shamir & Nemiroff (2005) to control the errors obtained in the posterior photometric analysis. In contrast to Press (1992) and Shamir & Nemiroff (2005), the proposed method quantizes not only the background pixels but also the foreground ones. Given that foreground values are typically 3 or 4 orders of magnitude higher than the quantization steps, foreground quantization can enhance compression performance while introducing arbitrarily small errors in the photometry analysis. The threshold t is used to classify pixels as foreground or background, while q_f and q_b are used to adjust the quantization coarseness for pixels likely to contain astronomical objects (i.e., the foreground) or not (i.e., the background). By carefully selecting these three parameters, it is possible to achieve the photometry fidelity requirements of the task at hand, while maximizing compression performance.

The classification of each pixel $I_{x,y}$ is stored in a binary mask, $\text{mask}_{x,y}$, with dimensions identical to those of the image. For each (x, y) position, $\text{mask}_{x,y} = 0$ indicates that the pixel is part of the background, and $\text{mask}_{x,y} = 1$ indicates that it is part of the foreground.

The quantization indices $I^q_{x,y}$ produced by the proposed flexible bi-region quantizer are defined using the classification

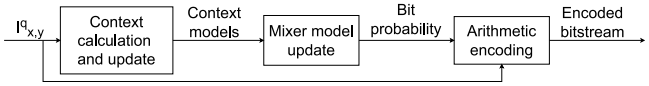

Figure 2. Main stages of the LPAQ9M entropy coder.

mask and the q_b and q_f parameters as

$$I^q_{x,y} = \begin{cases} \left\lfloor \frac{I_{x,y}}{q_b} \right\rfloor & \text{if mask}_{x,y} = 0 \\ \left\lfloor \frac{I_{x,y}}{q_f} \right\rfloor & \text{if mask}_{x,y} = 1 \end{cases}. \quad (5)$$

Finally, the encoder compresses the quantization indices without loss using LPAQ9M as described in Section 3.2. The classification mask is also losslessly compressed and transmitted as side information, which is typically 2–3 orders of size smaller than the rest of the compressed data.

On the decoder side, the quantization indices $I^q_{x,y}$ and classification mask, $\text{mask}_{x,y}$, are reconstructed using the LPAQ9M decoder. A lossy reconstruction of the original image pixels is then computed as

$$I^r_{x,y} = \begin{cases} \text{clip}((I^q_{x,y} + 0.5) \cdot q_b) & \text{if mask}_{x,y} = 0 \\ \text{clip}((I^q_{x,y} + 0.5) \cdot q_f) & \text{if mask}_{x,y} = 1 \end{cases}, \quad (6)$$

$\text{clip}(\cdot)$ is a function that rounds the argument to the closest integer value and truncates the result to the original pixel range, e.g., $[0, 2^{16} - 1]$ for an unsigned 16-bit image. Note that the constant addition of 0.5 in (6) results in the mid-point reconstruction of the samples, i.e., original samples that fall within a given quantization interval are reconstructed as the central value of the said interval.

3.2. LPAQ

In the proposed pipeline, lossless compression with LPAQ9M is applied to the quantization indices and the classification mask produced in the previous stage. LPAQ9M is a member of the PAQ family of lossless compressors, designed to have comparatively lower computational complexity than other members Mahoney (2002), Blaszczyk et al. (2012), Tang et al. (2022), Mahoney (2024). Its pipeline uses prediction and arithmetic coding, and combines multiple context models into a single probability distribution, which is used to drive an adaptive binary arithmetic encoder. The main stages of this algorithm are shown in Figure 2.

LPAQ9M uses seven context models to estimate the probability of each bit: five order-n models, $n \in \{1, 2, 3, 4, 6\}$, plus a word model and a match model. In the order-n models, the value of n input bytes and between 0 and 7 bits of the current sample are used to make probability estimations. Other members of the PAQ family use 16 order-n models plus several other types, which results in compression times 2 orders of magnitude longer than LPAQ9M. For astronomical images, other PAQ algorithms provide no significant compression performance advantage (Maireles-González et al. 2023), so they were not included in the proposed pipeline. Probabilities from different contexts are combined using an artificial neural network into a single probability estimation, which is used by the arithmetic encoder. After processing each bit, the probability distributions of all seven context models are updated, and so are the weights of the artificial neural network. These weights are updated considering that the prediction errors are the cost function to minimize.

The memory usage of the LPAQ9M encoder can be adjusted to approximately 2^N MB, where $N \in \{1, 2, \ldots, 9\}$ is a user-defined parameter. Smaller values of N result in lower computational complexity and faster execution times. On the other hand, larger values of N tend to produce better compression performance. For astronomical images, $N = 9$ is 2 orders of magnitude slower than $N = 2$, improving compression bit rates by less than 1%. Based on these observations, $N = 2$ offers the best trade-off for the task at hand, and is exclusively considered hereafter.

4. Experimental Results

The performance of the proposed compression pipeline is thoroughly analyzed in this section using a comprehensive set of five corpora comprising real astronomical image data from different telescopes. This analysis takes into account both the achieved compressed data volume reduction and the impact on the results of the photometric analysis. The test corpus and its properties are addressed in Section 4.1. A description of the compression and distortion metrics used in this study is provided in Section 4.2. The performance impact of the proposed pipeline's user-defined parameters is analyzed in Section 4.3, along with the problem with optimal parameter selection. Finally, a comparison of the proposed pipeline with the best-performing methods in the state of the art is available in Section 4.4.

4.1. Test Corpora

Experiments for this work have been performed on images from a representative selection of observatories around the world. Images have been gathered from telescopes of different diameters, preferable for different observation scenarios such as stars (isolated, open clusters, and globular clusters), galaxies (elliptical, spiral, irregular, face-on, edge-on, isolated, and interacting), and nebulae. The selected telescopes are the Isaac Newton Telescope (INT), the Jacobus Katpeyn Telescope (JKT), the Las Cumbres Observatory (LCO)—which groups the McDonald, Haleakala, Cerro Tololo, and Siding Springs observatories—, the Joan Orò Telescope (TJO), and

Table 1
Diameters of the Telescopes Sourcing Data for the Test Corpora

Telescope	Acronym	Diameter (m)
Isaac Newton Telescope	INT	2.5
Jacobus Katpeyn Telescope	JKT	1.0
Las Cumbres Observatory	LCO	0.4 and 2
Joan Orò Telescope	TJO	0.8
William Herschel Telescope	WHT	4.2

the William Herschel Telescope (WHT). The diameters of these telescopes are provided in Table 1.

A total of 206 astronomical images, 2.9 GB of uncompressed data, have been grouped into the five aforementioned corpora. Some basic properties of these images are summarized in Table 2, and additional information about the test corpora can be found in Maireles-González et al. (2023). All data are publicly available for download (Maireles Gonzalez et al. 2024).

4.2. Compression and Distortion Metrics

In this work, lossy compression methods are assessed by jointly considering the compressed data bit rate and the distortion introduced in photometric analysis. All photometric results are obtained with version 2.19.5 of SExtractor (Bertin & Arnouts 1996) configured to use a minimum threshold of 80 mag · arcsec^2, a saturation level of 60,000 Analogical to Digital Units, and a minimum pixel count of 9 above the threshold for object detection.

The compressed data bit rate indicates how many compressed bits are produced on average for each input sample, i.e., for each input pixel, as

$$\text{compressed data rate} = \frac{\text{Compressed data length (bits)}}{\text{Input image size (samples)}}. \quad (7)$$

It is expressed in bits per sample (bps), and lower values indicate more compression.

We study two aspects to assess the impact on photometry after lossy compression: the detection or not of each astronomical object, and the errors in the magnitude values assigned to each detected object. The ground truth of astronomical objects present in an image and their magnitude values is obtained by applying the SExtractor photometry software to the original images. The same software is applied to each image reconstructed after lossy compression and decompression, and the results are compared to those of the original. The following metrics are employed to compare these results:

1. *% recovered*. the percentage of recovered objects. It is calculated as the fraction of objects correctly detected from those detected in the original images. The criterion for a correct detection (a true positive) is for the two objects to have the same reported x and y position values with a maximum error of 0.5 pixels in each one.
2. *% recovered* <0.01 *mag*. the percentage of recovered objects with magnitude differences smaller than 0.01 mag. The 0.01 magnitude threshold has been described as the maximum that can be introduced without causing a significant disturbance in the scientific value of the data (Borucki & Lasher 2001; López-Morales 2006; Blake et al. 2008; McGraw et al. 2009; Bono et al. 2010; Di Cecco et al. 2010; Loh et al. 2012; Crawford 2017; Ivănescu et al. 2021).
3. *% recovered* < *measure error*: the percentage of recovered objects with magnitude differences smaller than the measurement error. This measurement error is provided by the photometry analysis software for each detected object. Magnitude differences smaller than this error indicate that the measurements in the reconstructed data are within the margin of error of the measurements in the original data.

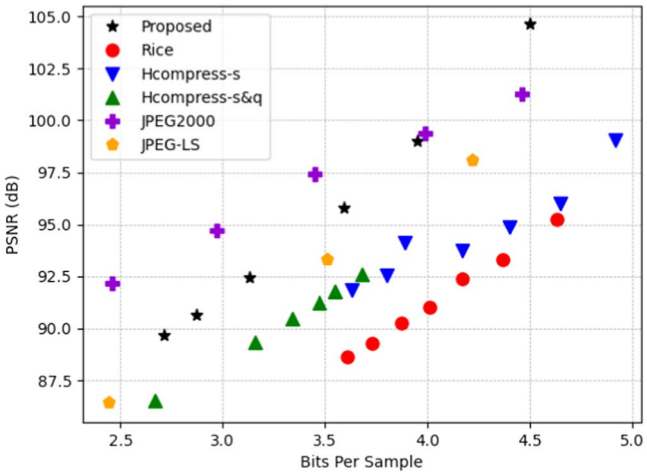

Figure 3. Average compressed data rate and Peak Signal-to-Noise Ratio (PSNR) rate-distortion over all test images.

Table 2
Properties of the Test Corpora

Telescope	Images	Dimensions	Average Entropy (bits)
INT	39	4200 × 2154	5.78
JKT	36	2100 × 2088	7.32
LCO	79	2112 × 3136	5.84
TJO	17	4096 × 4096	5.92
WHT	35	2501 × 2148	7.68

Note. Entropy results correspond to zero-order entropy.

Figure 3 shows the performance of the proposed tested techniques by measuring the distortion in the Peak Signal-to-Noise Ratio (PSNR). PSNR is a metric derived from the mean

square error (MSE) between the original and recovered image, calculated as:

$$\mathrm{PSNR} = 20\ \log_{10} \frac{2^B - 1}{\sqrt{\mathrm{MSE}(I_{x,y}, I^r_{x,y})}} \tag{8}$$

where B is the bit-depth of the images, for the current case $B = 16$. As can be seen, the proposed technique outperforms all fpack and JPEG-LS techniques for all bit rates, and JPEG 2000 for bit rates larger than 4 bps.

4.3. Optimal t, q_b, and q_f Value Selection

The values of the threshold t, background quantization step q_b, and foreground quantization step q_f of the proposed compression pipeline need to be adjusted to maximize performance according to the metrics defined in Section 4.2. An exploration of the parameter space has been conducted for each of the corpora described in Section 4.1, with the goal of finding the best configuration in each case.

The heuristic used to select the best combination of parameters t, q_b, and q_f for each corpus is as follows. The fixed requirement to maintain fidelity is to recover more than 99% of the original objects. Once this condition is satisfied, the optimal values are those having the lowest bps values, as shown in Figure 4. The results were tested for q_b and q_f ranging from 1 to 20, using parameter steps of 1. For the case of t, the parameter range was from the minimum pixel value found in each data set to a value 2000 counts higher, with steps of 100. Figure 4 shows different combinations of t, q_b, and q_f for each data set, to exemplify the compression and fidelity results behavior. The results are in pairs as one is for bps metric results and the other one is for % *recovered* metric results. As can be seen in Figure 4, usually $q_b \geqslant 6$ and $q_f \geqslant 6$ retrieve less than 99% of recovered objects and do not fulfill fidelity criteria. Usually $q_b = 3, 4, 5$ and/or $q_f = 4, 5, 7$ are the values typically selected as the best overall results, since they improve compression performance compared to lower quantization values and also fidelity. After an exhaustive search, the t values reported in Figure 4 are the ones with the minimum bps value for each q_b and q_f combination. The optimal compression values for each telescope are summarized in Table 3.

Figure 5 shows a visual comparison of a zoomed part of an image from the JKT data set. The pixel values of a small region of the original image are compared to the same pixel values obtained after lossy compression. The image is compressed at a rate of 3.5 bits per sample using the proposed technique, JPEG2000 and Hcompress, and the pixel differences are displayed. As can be observed, the differences in the proposed technique are smaller and smoother. The bi-region quantization is more suitable to adapt to the bi-modal nature of astronomical images (background and foreground), thus decreasing the pixel differences both near the sources (bottom left part of the difference matrix) and in the background regions (top right part of the difference matrix). The designed quantization scheme reduces the image entropy, improving compression while maintaining high photometric fidelity.

4.4. Comparison to the State of the Art

The performance of the proposed compression pipeline using the optimal parameters derived in Section 4.3 is compared to the state of the art in this section. As in the previous section, the figure of merit jointly considers the compressed data rate and the distortion introduced in the photometry. The representative selection of lossy compressors considered for comparison are described first. The results obtained are presented and discussed later.

4.4.1. Compared Algorithms

The de facto standard Fpack is evaluated in the two modes described in Section 2, Rice and Hcompress. Versions 1.7.0 and 3.470 of Fpack and CFITSIO are employed exclusively for this experiment. The values of the q and s parameters for Hcompress are selected for each data set as the ones that maximize compression performance while maintaining a percentage of recovered objects above 99%, as explained in the previous section. Fpack results have been tested for s and q ranging from 1 to 10, with parameter steps of 1. Hcompress is executed in a one-parameter mode using only s, henceforth called Hcompress-s, and also in a two-parameter mode using the same s value plus q, henceforth called Hcompress-s&q. Hcompress-s&q enhances the results of Hcompress-s, as will be seen next, and it has never been tested before in the literature. Unfortunately, the techniques described in Press (1992), Shamir & Nemiroff (2005), Pulido et al. (2020) could not be included in this comparison due to either the lack of source code, binaries or data sets available for public download, or the lack of a sufficiently detailed algorithm description to allow a full own implementation. Two more lossy compression techniques are included for completeness: JPEG 2000 (Taubman & Marcellin 2012), which uses a discrete wavelet transform, a bit-plane encoder and the MQ arithmetic coding (Baarir & Medouakh 2011); and JPEG-LS (Weinberger et al. 2000), which uses prediction, context modeling, and a Golomb entropy coding. These techniques have previously been tested in the lossless mode in Maireles-González et al. (2023) for this data set. JPEG 2000 Kakadu software (Taubman 2024) includes a lossy mode by adding a *rate* parameter, which indicates the bps value the compression performance reaches. In turn, JPEG-LS' near-lossless mode guarantees a maximum pixel (peak absolute) error m, only allowing integer values. The selected parameter values for each corpus and coding technique are shown in Table 4.

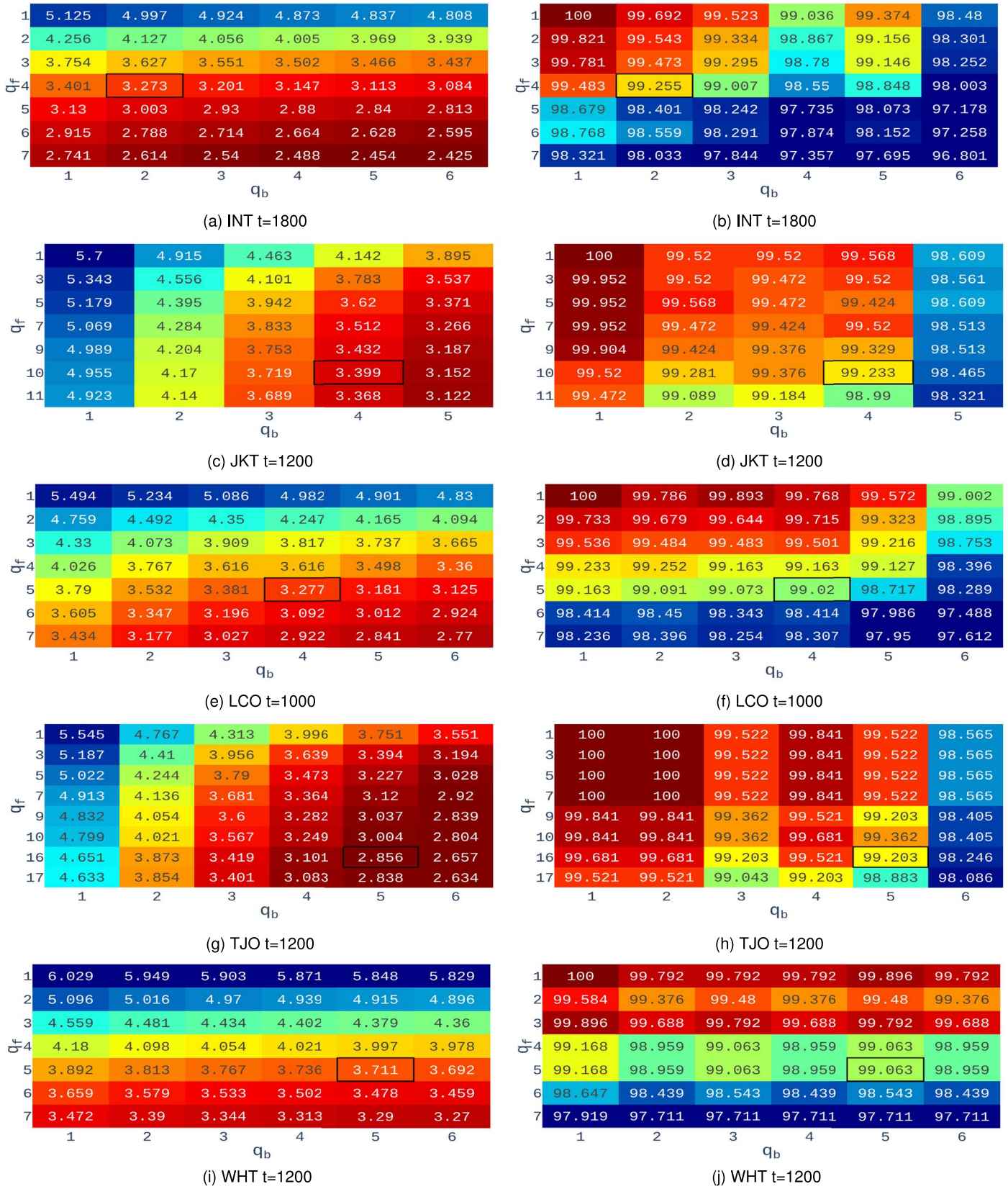

Figure 4. Heatmaps for each data set for the optimal t values. Red color indicating the best results. Left-side heatmaps representing the bps values and right-side heatmaps represent % recovered values. Best combination of parameters highlighted by a black box.

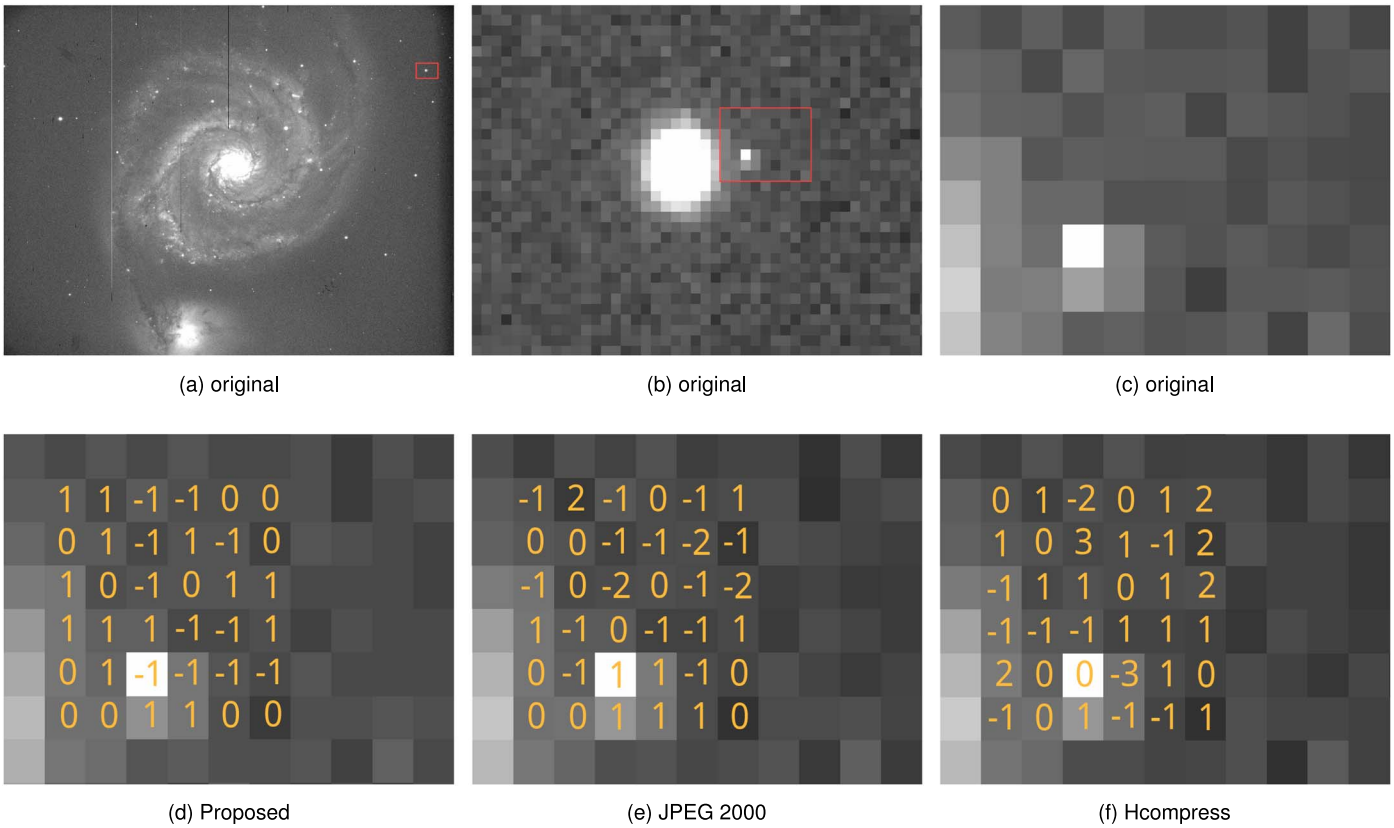

Figure 5. Visual comparison of the loss introduced by the main lossy techniques tested. The upper part shows a zoom of an M51 image included in the JKT data set. The lower part shows the zoomed region (c) after lossy compression performed by (d) the proposed technique, (e) JPEG2000 and (f) Hcompress. The pixel differences compared to the original image are shown above the pixels of the latter images.

Table 3
Optimal Parameters for the Proposed Method

Telescope	t	q_f	q_b
INT	1800	4	2
JKT	1200	10	4
LCO	1000	5	4
TJO	1200	16	5
WHT	1200	5	5

Table 4
Optimal Parameters for Rice, Hcompress, JPEG 2000 and JPEG-LS used in the Comparison

Telescope	Rice	Hcompress		JPEG2000	JPEG-LS
	q	s	q	Rate	m
INT	−7	−8	8	4	1
JKT	3	−8	4	4	1
LCO	−4	−8	4	4	1
TJO	−6	−8	4	4	1
WHT	−10	−8	4	4	1

4.4.2. Comparison Results

The compressed data rate and the percentage of detected objects are analyzed jointly. Figure 6 shows the average results for all data sets for our proposed pipeline and the algorithms in the state of the art. The 99% object detection level, i.e., the minimum acceptable fidelity, is highlighted in the figure for ease of reference. Tables 3 and 4 configurations are shown as the lowest bps points for each compression technique above the 99% fidelity level in Figure 6. Some additional points are also included to represent more and also less aggressive compression configurations and to better compare the state of the art techniques with the proposed one. As will be seen below, high bps results are of particular interest for satisfying restrictive photometric requirements. As can be observed, only the proposed pipeline and Hcompress-s&q offer the best compression when conservative quality levels are imposed. For the best compression scenario within the detection threshold of 99%, the results of the proposed pipeline indicate average improvements in the compressed data rate of 0.95 bps (24.05%), 1.09 bps (26.10%), and 1.72 bps (36.40%) over JPEG 2000,

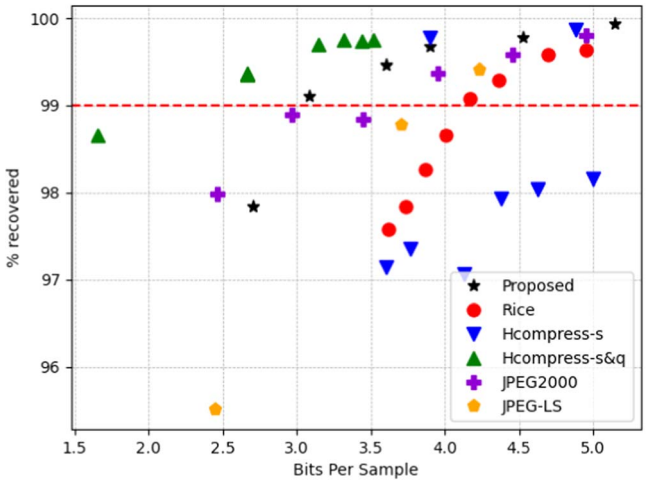

Figure 6. Average compressed data rate and percentage of detected objects over all test images.

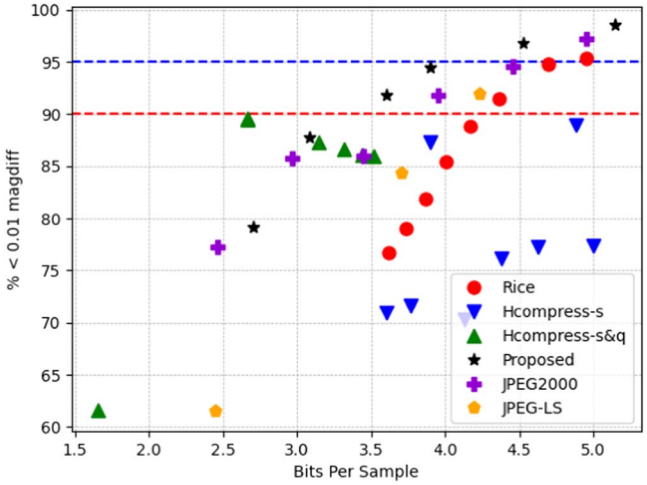

Figure 7. Average compressed data rate and percentage of detected objects with magnitude differences under 0.01 mag over all test images.

JPEG-LS and Rice, respectively. Hcompress-s&q best configuration has average improvements of 0.333 bps (11.10%) over the proposed pipeline. Conversely, at similar compressed data rates within the detection threshold of 99%, the proposed pipeline is able to preserve more relevant information for analysis in the images. The proposed pipeline recovers 0.59%, 0.30%, and 0.30% more objects compared to Rice, JPEG 2000, and JPEG-LS, respectively. Hcompress-s retrieves a similar recovery percentage, but, as shown next, the magnitude differences are notably higher. Hcompress-s&q has an improvement of 0.38% over the proposed method.

Even though the percentage of detected objects in the reconstructed images is informative, it is not enough to fully assess the fidelity of the photometry analysis: errors in the magnitude (intensity) of the detected objects are also critical. In Figure 7, the percentage of recovered objects with absolute magnitude differences lower than 0.01 mag (Borucki & Lasher 2001; López-Morales 2006; Blake et al. 2008; McGraw et al. 2009; Bono et al. 2010; Di Cecco et al. 2010; Loh et al. 2012; Crawford 2017; Ivănescu et al. 2021) are shown as a function of the compressed data rate. The 95% and 90% rates of objects with magnitude differences under 0.01 mag are highlighted in blue and red, respectively, in the figure for ease of reference. As can be observed, Hcompress results do not reach 90% at any bit rate. The proposed method reaches 90% value at 3.40 bps and 95% at 4.00 bps. Results indicate that, at a similar fidelity within the detection threshold of 99%, the proposed pipeline performs better compression than the state of the art methods. The results of the proposed pipeline indicate average improvements in the compressed data rate of 0.35 bps (8.86%), 0.63 bps (14.85%), and 0.77 bps (17.56%) over JPEG 2000, JEPG-LS, and Rice, respectively.

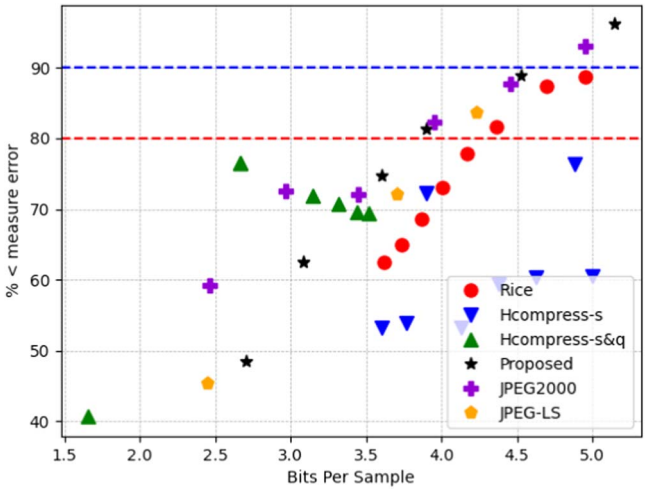

Figure 8. Average compressed data rate and percentage of detected objects with magnitude differences under measurement error.

It is also interesting to consider the estimated measurement error when performing photometry analysis. In addition to the position and magnitude of the detected objects, each magnitude measurement includes an absolute magnitude error margin. Results within this margin indicate that magnitude values produced using the compressed values after lossy compression are consistent with those obtained using the original, uncompressed data. In Figure 8, the percentage of recovered objects with absolute magnitude differences lower than the measure error is shown as a function of the compressed data rate. The 90% and 80% rate of objects with magnitude differences under 0.01 mag are highlighted in blue and red, respectively, in the figure for ease of reference. The proposed method reaches 80% value at 3.90 bps and 90% at 4.50 bps. In

Table 5
Average Lossless Compressed Data Rates in Bits Per Sample

Telescope	LPAQ9M	JPEG2000	Hcompress	Rice
INT	5.17	5.42	5.78	5.95
JKT	5.70	5.85	6.19	6.34
LCO	5.48	5.76	6.03	6.18
TJO	5.49	5.71	5.93	6.14
WHT	5.97	6.08	6.51	6.62
All	5.55	5.763	6.10	6.25

this bit range, its performance is similar to JPEG 2000, while it improves compression over JPEG-LS and Rice by 0.33 bps (7.73%) and 0.47 bps (10.67%), respectively, at a similar fidelity. Hcompress does not reach 80% value or higher at any bit rate.

Results indicate that the proposed pipeline is able to preserve high photometric fidelity in the reconstructed images, while significantly improving upon the compression performance of other lossy compression methods in the state of the art. It is important to stress that improvements over purely lossless compression methods are much larger, in spite of the comparable fidelity levels. Table 5 provides the best lossless compression data rates for astronomical images described in the literature (Maireles-González et al. 2023). Comparing these results to the ones obtained for our lossy proposed method, the compressed data set volume can be reduced up to 1.55 bps (38.75%) and 2.15 bps (27.94%) while maintaining 95% and 90%, respectively, of the detected objects with magnitude differences lower than 0.01 mag. For major photometric fidelity, data can be compressed up to 1 bps (18.93%) while maintaining 90% of the detected objects with magnitude differences lower than the photometric measure error. These improvements can be directly transferred to observatories' storage and transmission costs by adapting the proposed pipeline to the desired fidelity.

5. Conclusions

A new method to perform lossy compression that preserves the photometric properties and the number of sources of the original image is proposed. The technique is a combination of lossy bi-region quantization plus the LPAQ9M lossless compression technique with a specific parameter configuration. The bi-region quantization process parameters allow the pipeline to be adapted to the image characteristics, i.e., separating the natural bi-modal nature of the images (foreground and background) and achieving the desired compression or photometric fidelity requirements with high precision. A large and representative astronomical data set has been gathered to evaluate the lossy coding performance of relevant compression algorithms based on different coding paradigms. Compared to the results of the current compressor used in observatories, Fpack (Hcompress), the proposed method recovers 7.15% more sources with magnitude differences lower than 0.01 mag, and 9.13% more sources with magnitude differences lower than the photometric measure error, at a similar compression rate. Compared with the best lossless results, the proposed method can detect more than 99% of the objects in the original images, while the size of the compressed images is between 2.15 and 1.00 bps smaller than using lossless compression. For the first compression case, more than 90% of the recovered objects show magnitude differences of less than 0.01 mag; for the second one, more than 90% of them show magnitude differences equal to or lower than the magnitude measurement error of the photometry. This ensures the resulting recovered images are reliable in terms of photometric analysis. The proposed lossy compression pipeline allows the scientific community to significantly reduce data volumes without compromising the aforementioned analysis, which will help reduce storage and transmission costs for a rapidly growing volume of data archives.

Acknowledgments

This work makes use of observations from the telescopes operated on the island of La Palma by the Isaac Newton Group of Telescopes in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias, from Las Cumbres Observatory global telescope network and from The Joan Orò Telescope (TJO) of the Montsec Astronomical Observatory (OAdM), owned by the Catalan Government and operated by the Institute for Space Studies of Catalonia (IEEC). TJO data authorship is attributed to IEEC member Kike Herrero.

References

Baarir, Z. E., & Medouakh, S. 2011, International Journal of Computer Science Issues (IJCSI), 8, 531
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Blake, C. H., Bloom, J. S., Latham, D. W., et al. 2008, PASP, 120, 860
Blaszczyk, K., Rossmanith, P., Langer, D.-I. A., & Reidl, D.-I. F. 2012, PAQ Compression Algorithm
Bono, G., Stetson, P., Walker, A., et al. 2010, PASP, 122, 651
Borucki, W. J., & Lasher, L. E. 2001, in Third Workshop on Photometry (Palo Alto, CA: NASA), 1
Budding, E., & Demircan, O. 2007, Introduction to Astronomical Photometry, Vol. 6 (Cambridge: Cambridge Univ. Press)
Crawford, T. R. 2017, Photometric Accuracy/Precision - A Perspective, CTX, American Association of Variable Star Observers (AAVSO), [Online]. Available: https://www.aavso.org/photometric-accuracyprecision-perspective
Di Cecco, A., Becucci, R., Bono, G., et al. 2010, PASP, 122, 991
European Southern Observatory 2024, Data Reduction Frequently Asked Questions —How do I Uncompress .fz files?, [Online] https://www.eso.org/sci/data-processing/faq/how-do-i-uncompress-fz-files.html, 2024, accessed: 2024
Evans, M. S. 2010, SHPSA, 41, 86
Folk, M., Heber, G., Koziol, Q., Pourmal, E., & Robinson, D. 2011, in Proc. EDBT/ICDT 2011 Workshop on Array Databases, 36

Gailly, J.-l., & Adler, M. 1992, GNU Operating System (Free Software Foundation, Inc.)
Grange, Y. G., Pandey, V. N., Espinal, X., Di Maria, R., & Millar, A. 2022, arXiv:2202.01828
Hanisch, R. J., Farris, A., Greisen, E. W., et al. 2001, A&A, 376, 359
Henden, A. A., & Kaitchuck, R. H. 1982, Astronomical Photometry (New York: Van Nostrand Reinhold)
Isaac Newton Group Archive 2024, [Online] http://casu.ast.cam.ac.uk/casuadc/ingarch/, accessed: 2024
Ivănescu, L., Baibakov, K., O'Neill, N. T., Blanchet, J.-P., & Schulz, K.-H. 2021, AMT, 14, 6561
Kremer, J., Stensbo-Smidt, K., Gieseke, F., Pedersen, K. S., & Igel, C. 2017, in EEE Intelligent Systems, 32, 16
Las Cumbres Observatory 2024, Decompressing FPacked Data, [Online], https://lco.global/education/observing/fpack/, accessed: 2024
Loh, E. D., Biel, J. D., Davis, M. W., et al. 2012, PASP, 124, 343
López-Morales, M. 2006, PASP, 118, 716
Maireles-González, Ò., Bartrina-Rapesta, J., Hernández-Cabronero, M., & Serra-Sagristà, J. 2023, PASP, 135, 094502
Mahoney, M. V. 2002, The PAQ1 Data Compression Program Draft, Jan20
Mahoney, M. V. 2024, Data Compression Programs - LPAQ, [Online] http://mattmahoney.net/dc/, accessed: 2024
Maireles Gonzalez, O., Bartrina Rapesta, J., Hernández-Cabronero, M. ., & Serra Sagristà, J. 2024, Integer (L0) Astronomical Data from Several Observatories CORA.Repositori de Dades de Recerca (Universitat Autònoma de Barcelona), [Online]. Available: https://doi.org/10.34810/data1415
McGraw, J. T., Stubbs, C. W., Zimmer, P. C., Fraser, G. T., & Vivekanandan, J. 2009, Measurement Astrophysics (MAP) First Steps: A New Decade of Ground-based Photometric Precision and Accuracy in Astro2010: The Astronomy and Astrophysics Decadal Survey, Technology Development Papers, 34
Pence, W., Seaman, R., & White, R. 2021, arXiv:1112.2671
Pence, W. D., White, R., & Seaman, R. 2010, PASP, 122, 1065
Poudel, M., Sarode, R. P., Watanobe, Y., Mozgovoy, M., & Bhalla, S. 2022, ApSci, 2, 6202
Press, W. H. 1992, in A.S.P. Conf. Series, 25, Astronomical Data Analysis Software and Systems I, ed. D. M. Worrall, C. Biemesderfer, & J. Barnes, 3
Pulido, J., Zheng, C., Thorman, P., & Hamann, B. 2020, MNRAS, 493, 2545
Rice, R., Yeh, P.-S., & Miller, W. 1993, in 9th Computing in Aerospace Conf., 4541
Romanishin, W. 2006, An Introduction to Astronomical Photometry Using CCDs, Vol. 17 (Univ. of Oklahoma)
Shamir, L., & Nemiroff, R. J. 2005, AJ, 129, 539
S. T. S. Institute 2024, Hubble Space Telescope Publication Statistics 1991-2024, [Online] https://archive.stsci.edu/hst/bibliography/pubstat.html, accessed: 2024
Tang, D., Sun, X., Guan, N., Kuo, T.-W., & Xue, C. J. 2022, in Int. Conf. on Field-Programmable Technology (ICFPT) (Piscataway, NJ: IEEE), 1
Taubman, D. 2024, kakadu software, [Online] https://kakadusoftware.com/, accessed: 2024
Taubman, D., & Marcellin, M. 2012, JPEG2000 Image Compression Fundamentals, Standards and Practice, Vol. 642 (Berlin: Springer)
Tody, D. 1993, in A.S.P. Conf. Series, 52, Astronomical Data Analysis Software and Systems II, ed. R. J. Hanisch, R. J. V. Brissenden, & J. Barnes, 173
Weinberger, M. J., Seroussi, G., & Sapiro, G. 2000, ITIP, 9, 1309
White, R., Postman, M., & Lattanzi, M. 1992, Digitised Optical Sky Surveys (Berlin: Springer), 167