Compression algorithm for Multi Element Telescope for Imaging and Spectroscopy (METIS)

Compression algorithm for Multi Element Telescope for Imaging and Spectroscopy (METIS)

REVIEW OF CCSDS 123 STANDARD

The algorithm presented in CCSDS 123-0 standard [6] has been defined by the MHDC working group of the Consultative Committee for Space Data Systems, composed by world major space agencies, aiming to provide a standard for on-board lossless coding of data coming from multi- and hyperspectral imagers and sounders. It achieves state-of-the-art lossless compression performance for a wide set of imagers collected by heterogeneous sensors, requiring low computational resources.
The standard follows the typical scheme of lossless compression, employing a prediction-based approach; this aspect provides a further advantage: as opposed to the transform-based scheme, it guarantees a low computational complexity that is a key constraint to be suitable for on-board applications.
The input to the compressor is a three-dimensional array of integer sample values; the variable-length compressed output is an encoded bitstream from which the input image can be exactly recovered.
The compression process is simply composed by two functional parts: a predictor followed by an encoder. The prediction of the sample $s_{x,y,z}$ consists in the calculation of a predicted value $\hat{s}_{x,y,z}$ from which the mapped residual $\delta_{x,y,z}$ can be derived.
The procedure to compute the predicted value involves a three-dimensional neighborhood: first, the predictor computes a local average of neighboring sample values; a weighted sum of neighboring samples from the current and the user-defined number of previous bands is then calculated, and finally the predicted value can be obtained by subtracting from this weighted sum, the previously computed local average.
One of the main features of this standard is the capability of the predictor to adapt to the local statistics, using the prediction error of every pixel to enhance the computation of the following predicted values, through a weight update.
The mapped residual $\delta_{x,y,z}$, of every pixel, a non-negative integer value, is encoded using a variable-length binary codeword composing the body of the compressed file. The variable-length codes are adaptively selected based on statistics that are updated after each sample is encoded; separate statistics are kept for each spectral band and the compressed image size does not depend on the order in which residuals are encoded. Upon the body of the compressed image, there is the header holding all the information of the image essential for decoding and to reconstruct exactly the input image.

PROPOSED ALGORITHM

The compression algorithm we are proposing for METIS consists on an integration and adaptation of the released standard described above, to meet the mission target: a flexible compressor has been studied ad-hoc for the nature of the images that are expected to be acquired by the sensors. Two key aspects that have been considered for the purpose are respectively the geometry of the acquired images and the usually slow temporal variability of those kind of acquisitions, since the solar phenomena take place once in a while and concern only a specific section of the images, leaving most pixels almost unaltered or with small variations. It is tailored to the simulated sensed images, and it allows the user to choose among a wide set of different combinations of parameters to compress more, or to keep the acquisitions coded losslessly if for example a Coronal Mass Ejection is ongoing, or even to select some regions to be kept with a higher degree of quality with respect to the rest of the image.

From Multi- Hyperspectral To Multi-temporal Image Data

The predictor of the CCSDS 123 standard employs a three-dimensional neighborhood to compute the predicted sample value, as is shown in Fig 1, analyzing pixels from preceding bands of a multispectral image; for our purpose, we investigated the possibility to switch the third dimension, from wavelength to time: a way to exploit the similarity of multiple consecutive acquisitions is to combine them into a three-dimensional cube that is used to feed the compressor.

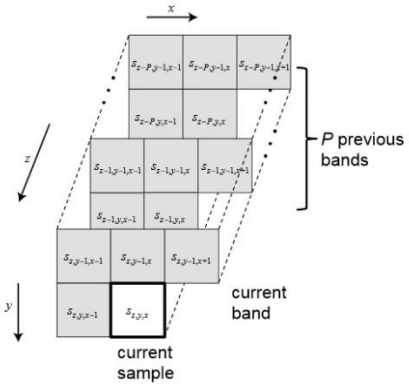

Fig 1: Typical CCSDS 123 prediction neighborhood

In this way, temporal correlation would enhance the calculation of predicted pixels values, allowing to take advantage of the efficiency of the 3D images compression of the standard in an alternative way, thereby increasing performance.
A possible drawback lies in the need of storing on the on-board memory a given number of acquisitions before compressing and send them down to the ground segment; actually the storage capacity could be limited on a space-craft; nevertheless, simulations will show that a limited number of consecutive images, in the order of few units, is sufficient to reach the top performance, and the gain in terms of bitrate or compression ratio justifies an additional storage capacity if needed.

Insertion Of Near-Lossless Routine

While the multi-temporal routine is a simple adaptation, changing only the approach to the compression and the construction of the 3-D data cube, a consistent change in the algorithm structure is entailed by the insertion of a near-lossless compression modality.
The expected images have different levels of priority in different regions: apart from the coronal area, nearest to the sun, the rest of the acquisition can tolerate a certain degree of error, user-defined and within specific bounds, leading to a potential increase of the compression ratio. This is why a variable-quality lossy routine has been added to the compression scheme.

The near-lossless routine has been designed through the insertion in the prediction loop of a uniform scalar quantizer, as described in Fig 2, and adding a local decoder, in such a way that the locally recovered pixel values, and not the original ones, are feeding the prediction scheme, to allow a correct decoding of the compressed stream at the receiver side.
The prediction residual, before feeding the entropy coder is quantized as:

$$e^{q}_{z,y,x} = sign(e_{z,y,x}) \cdot round\left(\left(|e_{z,y,x}| + \frac{Q-1}{2}\right)/Q\right) \quad (1)$$

being Q an odd integer quantization step size, obtained by δ, user defined positive integer, as $Q = 2\delta + 1$. The mapping and the encoding of the quantized residual stays the same as it is for the standard; while the de-quantized value, obtained by simply multiplying Q and the value in (1), is added to the predicted value and the summation, restoring the recovered value with an error bounded in $[-\delta,+\delta]$. This value is stored in memory and used for the prediction of successive samples. So a further structure is necessary to store the restored pixels values.

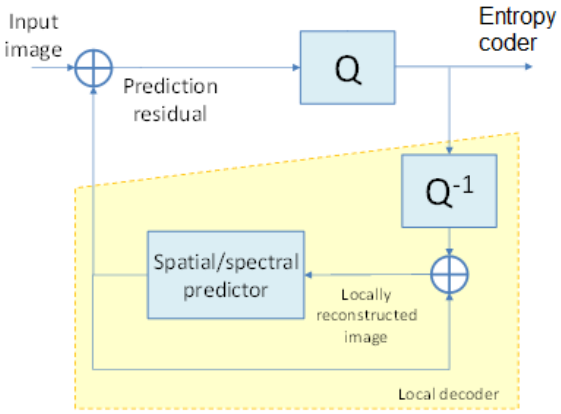

Fig 2 - Modified prediction loop

Radialization

Analyzing the geometry of the expected solar acquisition, the development of the radialization routine leads to a more efficient use of the predictive algorithm and a number of other advantages for the processing of the sensed images; it consists of a pre-processing of the acquisitions before feeding the compression algorithm, simply rearranging the pixels in a more appropriate geometry to exploit properties such as radial symmetry. It can be considered as a change of reference system, from a Cartesian to a polar one; for each pixel, the distance from and the angle with respect to the image center is computed; each pixel will belong to the row corresponding to its distance to the center, and the relative angles are ordered for every row, in order to keep a proper sorting for all the pixels with the same distance, so belonging to the same circular corona. After ordering the angles, the radialized image is shifted to the center to maximize correlation among neighboring pixels. A pseudo-code of the algorithm is presented below in Algorithm 1.

Algorithm 1 - pseudocode of radialization routine

$[M, N] = size(input\_image)$
for $x = 0; x < M; x++$ ***do***
  for $y = 0; y < N; y++$ ***do***
    $dist(x,y) = \langle\sqrt{(x - x_c)^2 + (y - y_c)^2}\rangle$
    if $R \leq dist \leq r$ ***do***
      $radial\big(dist, pos(dist)\big) = input\_image(i,j)$
      $angles\big(dist, pos(dist)\big) = 2\ \tan^{-1}((y - y_c)/(\sqrt{(x - x_c)^2 + (y - y_c)^2} + x - x_c))$
    end if
    $pos(dist)++$
  end for
end for
$ordered\_angles(x,y) = \boldsymbol{sort}(angles, \text{rows})$
for $x = 0; x < M; x++$ ***do***
  for $y = 0; y < N; y++$ ***do***
    $K \leftarrow ordered\_angles(X,k) == angles(X,y)$
    $ordered\_radial(x,y) = radial(x, ordered\_angles(x,y))$
  end for
end for
for $x = 0; x < M; x++$ ***do***
  for $y = 0; y < N; y++$ ***do***
    $shift\ ordered\_radial(x,y) \rightarrow ordered\_radial(x, N - pos(x) + y)$
  end for
end for

A further functionality of this routine is the possibility to specify internal and external radii, to apply a masking on the acquisitions, if a user desires to avoid coding the central occulter part or the external angles. An example of the radialization image processing is shown in Fig 3: it is one of the STEREO acquisitions, used for tests, and the figure depicts the original image, then the masked image, then the radialized image and finally the centered image, which is used as input to the compression algorithm.

A key aspect of this approach is that, differently from the common procedure when changing a reference system, no interpolation is performed among pixels, so that it is possible to keep the whole compression process lossless. Then, since radialization is a simple rearranging of pixels, maps can be derived in a way that the whole algorithm is executed once, and images can be radialized by simply performing the mapping with the obtained permutation structures; this leads to a really computationally light procedure.
Indeed, radialization increases correlation among neighboring pixels, enhancing the prediction and consequently the algorithm compression performance; moreover, it makes very easy to identify areas of interest: rows in the radialized image corresponds to circular coronae in the original acquisitions, and if an user desires to keep the inner part of the solar corona coded losslessly while tolerating a certain degree of error in the rest of the image, it is possible to employ a variable-quality quantization accordingly.

This kind of implementation brought as well some drawbacks: the radialized structure has not a rectangular shape, which is why zero padding values have been added in order to be able to feed the compressor with a rectangular image; this may lead to an increase in requested storage size with respect to the original acquisition and may decrease the effectiveness of the predictor adaptation, because of continuous re-adaptations caused by moving from padding to effective pixels values regions. The solution is simply storing an extra array, where the effective length of each radialized image row is kept, and further adapting consequently the predictor, in order to discern between actual pixels and padding ones, predicting and coding the former only, and skipping the latter; this leads to a significant increase in prediction performance but also lightens the compressor from a computational point of view.

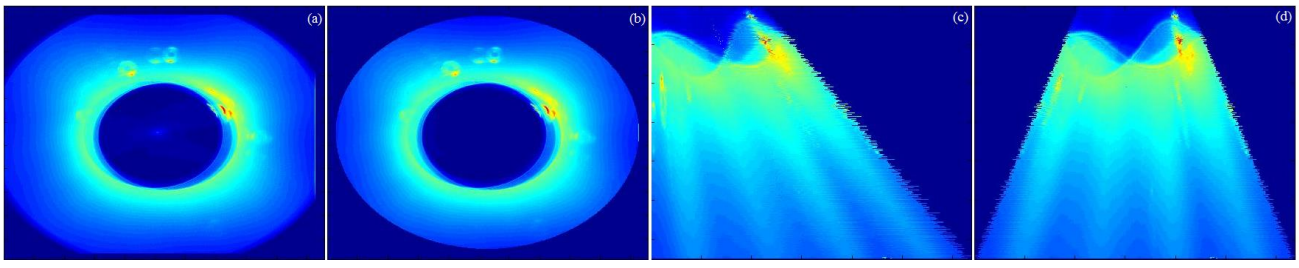

Fig 3 - Radialization sequence

Further Optimizations

The standard itself allows users to choose among a wide range of different options, in order to be robust with a large set of imagers; for the mission purposes, some of them have been discarded, such as the block-adaptive entropy coding routine. Another way to simplify the software structure is the choice of a single prediction mode, i.e. the neighbor-oriented, optimized by radialization, rather than column-oriented mode, which performs well in presence of striping noise, a case not applicable in this context. The last pre-processing routine that has been added to the software is the binning, applicable both on the original and radialized image; this is another advantage of applying radialization, namely the possibility to have a radial binning on the acquired image.
In conclusion the developed algorithm presents a wide set of ad-hoc options that can be even combined together to reach the desired target in compression keeping the quality for specific areas unaltered depending on user requirements.

PERFORMANCE ASSESSMENT

The image already mentioned and shown before, belongs to a set of acquisitions that have been used to test the effectiveness of the algorithm, since their nature looks similar to the expected images from METIS; they come from STEREO (Solar TErrestrial RElation Observatory), a mission consisting in the launch of two nearly identical space-crafts, one ahead of Earth in its orbit, the other trailing behind, enabling a first-ever stereoscopic imaging of the Sun and solar phenomena, as CMEs (Coronal Mass Ejections). An occulter conceals the Sun center, to highlight the real section of interest that is the Solar corona, in particular the area closer to the Sun. The size in pixels is 1088 × 1056, so a circular section would not impact in losing useful pixels. A three-dimensional cube of six successive acquisitions in time has been used to feed the algorithm; since the size of the input image varies when switching from original to radialized image, the results presented here are evaluated by the *compression ratio with respect to the original (not radialized) image size*, which is more fair and appropriate in this case, that is 2,19 MB for a single acquisition, and hence 13,15 MB for the whole cube. The radialized image size is 18,22 MB including padding zeros.

Compression Results

Some tests have been run to check algorithm performance with different combinations of the mentioned parameters, assessing the performance of radialization, while still keeping some portions of the images coded losslessly. The tests have been run on both STEREO coronagraphs acquisitions, constituting 3-D cubes of original and radialized images; the radialized images are obtained with internal radius equal to zero, so without discarding any pixels of the central part of the acquisitions, and an external radius of 500, as is shown as an example in Fig 4 and Fig 5.
The near-lossless mode, extended with the variable-quality quantization routine, has been run on the radialized images, applying lossless compression only to the most interesting portion of the image from a scientific point of view; it has been located approximately between rows 175 and 300. Table 1 summarizes the quantization parameter for the image rows in the three variable-quality quantization paths that are used in the experiment, which will be denoted as VQ1, VQ2 and VQ3 respectively.
Fig 6 shows the correspondence between rows selection in the radialized image and the selection of a coronal-shaped area in the original acquisition. It is clear how it becomes much simpler to select coronal-shaped areas as regions of interest thanks to radialization.
Tests have been run trying to vary the number of successive acquisitions to be used for prediction for both cases, original and radialized.
Fig 7 shows the better performance of applying radialization to acquisitions, and further how it allows to reach top performance by using a small value of consecutive acquisitions in the prediction loop, which could be an advantage if there are on-board memory constraints. Indeed, radialized images compression reaches the best compression ratio using only 4 consecutive acquisitions, while the compression of the original images seems works better using all the consecutive images for prediction; moreover, the compressed stream is clearly smaller when applying radialization, which means that this routine allows to compress much more compared to compressing the original acquisitions.

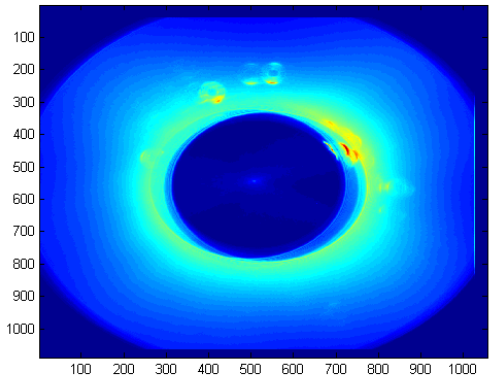

Fig 4 - Original image

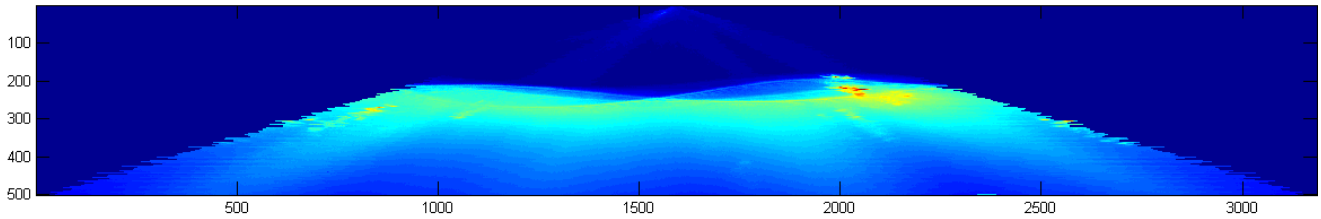

Fig 5 - Radialized image

Table 1 - Variable quantization paths

Rows	**VQ1**	**VQ2**	**VQ3**
0-175	5	10	25
175-300	0 – lossless	0 – lossless	0 – lossless
300-500	1	5	10

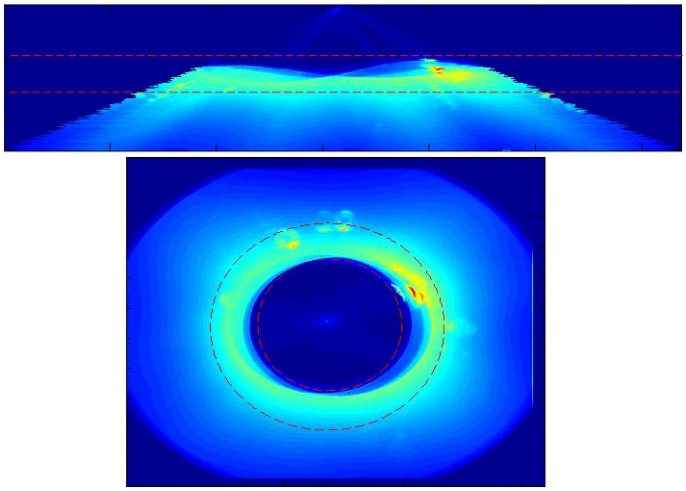

Fig 6 - Area selection in radialized and original acquisition

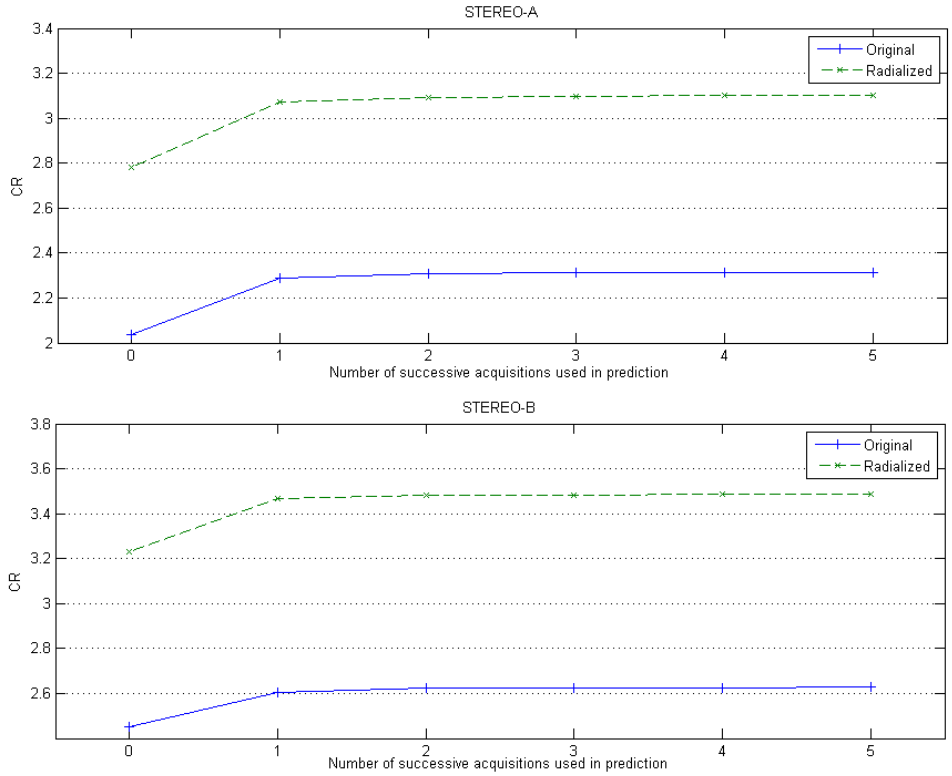

Fig 7 - Original and radialized image compression varying number of consecutive acquisitions used in prediction

Table 2 compares the performance of the two cases, showing the compression ratio computed as explained before, fixing the number of previous acquisitions at 4, and varying other parameters, in particular the quantization one. The variable-quality quantization task has been tested only for radialized case, since it does not make much sense to select rectangular areas from the original acquisitions, while it is possible to losslessly code the inner solar corona for the three cases specified in Table 1, which allow to select a circular corona in the original acquisition, and then varying the degree of quantization for the other areas. Finally, a 2 × 2 binning for both cases has been tested as well.

Table 2- Compressor performance

	STEREO-A		**STEREO-B**	
	Original	Radialized	Original	Radialized
δ = 1	2.9951	3.9224	3.5351	4.5589
δ = 2	3.4617	4.4739	4.1991	5.3177
δ = 5	4.4739	5.6818	5.7062	7.0539
δ = 10	5.7538	7.2104	7.8002	9.4327
VQ1	-	3.9051	-	4.5352
VQ2	-	5.5181	-	6.8039
VQ3	-	6.8706	-	8.8610
2x2 binning	10.0179	13.5015	11.6557	15.4900

Results show explicitly that radialization leads to a significant gain in compression performance; the gap becomes even higher when near-lossless compression is adopted, or binning as well. So in addition to a facilitated handling of the image processing for the mission purposes, radialization is capable to optimize and to enhance the compression task.

CONCLUSIONS AND FUTURE WORK

In this paper, we have shown our work, proposing an algorithm based on a standard and suitable for on-board application since it is computationally light; it is designed ad-hoc for this kind of acquisitions and in general for Solar images, exploiting their structure and radial geometry. It allows the user to select coronal-shaped areas of interest by simply varying some input parameters, to mask the image as desired or even to apply different kinds of binning to further reduce the data volume, achieving consistent performance in any case.
Some final optimizations are still ongoing, and particularly a radial binning: still exploiting the radialized image, it will be possible to apply more than one type of binning within the same image, as it is for variable quantization; in this way, going back to the original acquisitions, it will be possible to have coronal-shaped regions of the image in which different binning levels are applied. All these options can be combined together to increase the compression ratio.

ACKNOLEDGEMENTS

This work is supported by the Italian Space Agency under contract number I/043/10/0. The authors would like to thank the METIS team and the industrial partners for their invaluable effort in providing technical and managerial support.

REFERENCES

[1] S. Fineschi, E. Antonucci, G. Naletto, M. Romoli, D. Spadaro et al. "METIS: a novel coronagraph design for the Solar Orbiter mission", *Proc. SPIE 8443,* Space Telescopes and Instrumentation 2012: Ultraviolet to Gamma Ray, 84433H (September 7, 2012)

[2] A. Bemporad, G. Nicolini, B. Zhao, E. Magli, Performance assessment of lossy compression algorithms for application to METIS solar probe, ESA, 2012 Onboard Payload Data Compression Workshop (OBPDC), Barcelona (Spain) Oct. 2012, pp. 8, 2012

[3] Department of Information and Communications Engineering, Universitat Aut`onoma Barcelona, "IDC user manual (version 2.0)"

[4] Majid Rabbani*, Rajan Joshi, "An overview of the JPEG2000 still image compression standard" Signal Processing: Image Communication 17 (2002) 3–48

[5] Shantanu D. Rane and Guillermo Sapiro, Member, IEEE, "Evaluation of JPEG-LS, the New Lossless and Controlled-Lossy Still Image Compression Standard, for Compression of High-Resolution Elevation Data" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 39, NO. 10, OCTOBER 2001

[6] "Lossless Multispectral & Hyperspectral Image Compression. Blue Book. Issue 1. May 2012, http://public.ccsds.org/publications/archive/123x0b1ec1.pdf," [Online].