The Compression Algorithm of Hyperspectral Space Images Using Pre-byte Processing and Intra-bands Correlation

Alexander Zamyatin
Institute of Cybernetics of National Research
Tomsk Polytechnic University
Tomsk City, Russia
e-mail: zamyatin@tpu.ru

Assiya Sarinova
Innovation Eurasia University
Pavlodar City, Kazakhstan
e-mail: assiya_prog@mail.ru

Pedro Cabral
Institute of Statistics and Information Management
New University of Lisbon
Lisbon, Portugal
e-mail: pcabral@isegi.unl.pt

Abstract — **This paper concerns the compression of hyperspectral data from Earth remote sensing by suggesting a multistage algorithm to increase the compression ratio, using a formation of auxiliary data of the large redundancy in their byte presentation, and taking in account the correlation intra-bands. Here are presented the results of the studies on the effectiveness of compression of hyperspectral space images made by the proposed compression algorithm with the universal and specialized compression software.**

Keywords - remote sensing; hyperspectral images; lossless compression; intra-bands correlation; byte representation of data.

I. INTRODUCTION

Modern centers for space monitoring and systems of Earth Remote Sensing (RS) continually process, archive and distribute data which constitute tens and hundreds of gigabytes [1-9]. A key problem in the process is compressing RS data to increase effectiveness of a data transfer via connection channels of limited carrying capacity and archiving in RS storage subsystems of a limited capacity. The necessary classification of this data must be of the highest value. That is why it is more appropriate lossless compression, which is free of any distortions of statistic brightness characteristics of restored data.

The solution of the compression problem, which is the most accessible for practical implementation presupposes the usage of universal and widely-known algorithms and means of compression, for instance, in the archival software *WinRar*, *WinZip* or compressor *Lossless JPEG (JPEG-LS)* on the base of *JPEG* [10-15] image compression standard. However, RS data is classified by various characteristics – spectral, radiometric, spatial resolutions and by geometrical size of the scene. The above-mentioned universal means of compression fail to consistently consider the variable differences [16-19]. Thus, there are multispectral and hyperspectral Aerospace Images (AI), which have essentially different parameters of spectral resolution and are characterized by dependence (correlation) between data of different bands [20]. If for multispectral aerospace images the coefficient of mutual correlation of bands is $R \in (0.3; 0.8)$, then for hyperspectral AI representing values of brightness which are received in various spectral bands with the high spectral resolution, the correlation of neighboring bands is $R \approx 1.0$. This demonstrates high redundancy of data and expediency of applying this feature while compression [21][22].

Besides, knowing the correlation value (intra-bands correlation) between bands of hyperspectral AI, it is expedient to operate values of deviations (difference) between it and actual reference values, and that will allow to reduce the range of data change and hence will demand a smaller number of categories for their storage. The effective usage of specialized means of the compression is possible, considering the above-stated features of hyperspectral AI.

The software intended for the analysis of spatial data and processing of AI are *ERDAS Imagine*, *ERDASER Mapper*, *ArcView GIS*, *GeoExpress* and others often dispose of specific modules for compression of AI. Thus, the package *ERDAS Imagine* utilizes the compression tool for images in the format of MrSID (*IMAGINE MrSID Desktop Encoder* and *IMAGINE MrSID Workstation Encoder*), based on wavelets and intended for lossy compression of large RS images. The system *ERDASER Mapper* has modules of loss compression on the basis of the standards *JPEG2000* and *ECW* [23]. In the package *ArcView GIS* there is a module *MrSID* which compresses raster images files with loss [24]. The software product *GeoExpress* is intended for compressing raster data with the use of popular formats *MrSID* and *JPEG2000* [25]. Thus, all commercial systems of RS data processing are have all means of loss compression on the basis of well-known standards.

Today hyperspectral RS images lossless compression research attempt to apply various approaches and methods [26-34]. The separate stages of transforming data while compressing AI and possibility of decreasing power inputs and algorithmic complexity are discussed. Additionally, there are various attempts at adapting standards which have successfully proved for compressing usual images for compressing hyperspectral AI.

 ISBN: 978-1-61208-326-1

II. DESCRIPTION OF THE ALGORITHM

However, not all details of the original algorithms of compression are clear. Their numbers exceed the possibilities of most widespread means of compression; when applied to hyperspectral AI with various characteristics is uncertain.

It is our purpose to promote further search for approaches to lossless compression of hyperspectral RS images, substantially free from the disadvantages of existing universal and specialized means of compression.

Considering the features of hyperspectral AI and some details of existing analogues, the most expedient solution of the problem of compressing hyperspectral AI is by multi-stage transformations: first the advantages of universal traditional approaches for data compression, and secondly to consider the specificity of hyperspectral data. The algorithm embodying this approach and some of its results is discussed below.

Considering the specificity of compressing hyperspectral AI, the proposed algorithm has the following stages:

1. To consider the functional dependence of values of brightness (albedo) between various bands of images, by calculating the correlation and of deviations (differences) of initial data and the values of those found for functional dependence.
2. Creation of auxiliary structure of data on the basis of the initial hyperspectral AI, storing the unique pair groups of values of elements in a byte representation, and addressing references on these unique pair groups as well.
3. Compression of received data transformations with standard entropy algorithm by processing the generated auxiliary structures of the data.

Let us consider the details of the above-mentioned stages.

In the first stage, the value of deviations of the linear dependence on the matrix of values $\mathbf{I}[m,n,k]$.

For step-by-step description of the first stage consisting in searching correlation and deviations it is necessary to account the following objects:

- An initial image – the matrix of values of the image $\mathbf{I}[m,n,k]$, where m,n,k – are indices of the lines, columns and bands of the initial image, $m = 1,2,\ldots,M$, $n = 1,2,\ldots,N$, $k = 1,2,\ldots,K$;
- $\boldsymbol{R}[k]$ – a file for "level-by-level" preservation values of correlation R between the neighboring bands (layers);
- $\mathbf{Q}[k]$ – a file for placing values of the mathematical expectation for each band $\mathbf{I}[m,n,k]$;
- $\mathbf{L}[k]$ – a file for preserving the linear dependence;
- $\mathbf{I}'[m,n,k]$ c a file for placing values of differences (deviations) between $\mathbf{L}[k]$ and $\mathbf{I}[m,n,k]$.

Step 1. To calculate a mathematical expectation m_k of each band of the initial image $\mathbf{I}[m,n,k]$ and to place values to the file $\mathbf{Q}[k]$, as in (1):

$$\mathbf{Q}[k] = \sum_{k=1}^{K} \mathbf{I}[m,n,k] \times \rho_{\mathbf{I}}^{k} \tag{1}$$

where $\rho_{\mathbf{I}}^{k}$ – relative frequency of occurrence of values of the image $\mathbf{I}$, k=1,2,…, K.

Step 2. To calculate (on the base of $\mathbf{Q}[k]$ and $\mathbf{I}[m,n,k]$) correlation R for each pair of all available K bands of the initial image $\mathbf{I}[m,n,k]$, as in (2):

$$\mathbf{R}[k] = \frac{\sum_{m,n} (\mathbf{I}[m,n,k] - \mathbf{Q}[k])}{\sqrt{\sum_{m,n} (\mathbf{I}[m,n,k] - \mathbf{Q}[k])^2}} \tag{2}$$

To place the result to $\mathbf{R}[k]$, k=1,2,…, K-1.

Step 3. To calculate (on the base of $\mathbf{Q}[k]$ and $\mathbf{R}[k]$) a linear dependence of the kind $L=m_k \times R$ for each pair of the available K bands of the initial image $\mathbf{I}[m,n,k]$. Result should be placed to $\mathbf{L}[k]$, $\mathbf{L}[k] = \mathbf{Q}[k]\times\mathbf{R}[k]$, k=1,2,…,K-1.

Step 4. To calculate (on the base of $\mathbf{R}[k]$ and $\mathbf{L}[k]$) the difference between elements $\mathbf{L}[k]$ and corresponding values of the initial image $\mathbf{I}[m,n,k]$ in each of K bands, as in (3):

$$\mathbf{I}'[m,n,k] = \begin{cases} by\ \mathbf{R}[k] > 0, \mathbf{I}[m,n,k] - \mathbf{L}[k-1] \\ by\ \mathbf{R}[k] \le 0, \mathbf{I}[m,n,k] \end{cases} \tag{3}$$

for m=1..M, n=1..N. The result is placed to $\mathbf{I}'[m,n,k]$.

Step 5. To transform the negative values $\mathbf{I}'[m,n,k]$ to positive ones in the byte representation as numbers with a sign demand more bytes than those without a sign (4):

$$\mathbf{I}'[m,n,k] = \begin{cases} \text{at } \mathbf{I}[m,n,k] \ge 0,\ 2\times\mathbf{I}[m,n,k] \\ \text{at } \mathbf{I}[m,n,k] = 0, -2\times\mathbf{I}[m,n,k] - 1 \end{cases} \tag{4}$$

The result is the file $\mathbf{I}'[m,n,k]$ considering values of intra-bands correlation $\boldsymbol{R}[k]$.

The essence of the second stage consists of forming a file of unique pair groups of values which represent the initial image in the byte representation. Then, the file is formed containing references to the same pair group of values.

The algorithm proceeds with two additional objects:

- $\mathbf{M}[j,k]$ – a file of unique pair groups of values of the initial image in the byte representation;
- $\mathbf{D}[j,k]$ – a file for entering references (to unique pair groups of values).

In second stage of transformation, the step-by-step elaboration is shown in Figure 1:

Step 1. To form (on the base of **I'**[*m*,*n*,*k*]) a file **M**[*j*,*k*] for each band *K* adding unique pair groups of values in the byte representation from the file **I'**[*m*,*n*,*k*]. To place the result to **M**[*j*,*k*], *j*= 1,2,..,*J*. If repeated pair groups of values are absent, then $J = (M \times N \times K)/2$, *k*= 1,2,..,*K*.

Step 2. To form (on the base of **M**[*j*,*k*]) a file **D** [*j*,*k*], putting down references to unique pair groups of values from the file **M**[*j*,*k*] to **D**[*j*,*k*]. To place the result to **D**[*j*,*k*], $j<J$ as $j = M \times N \times K$, *k*= 1,2,..,*K*.

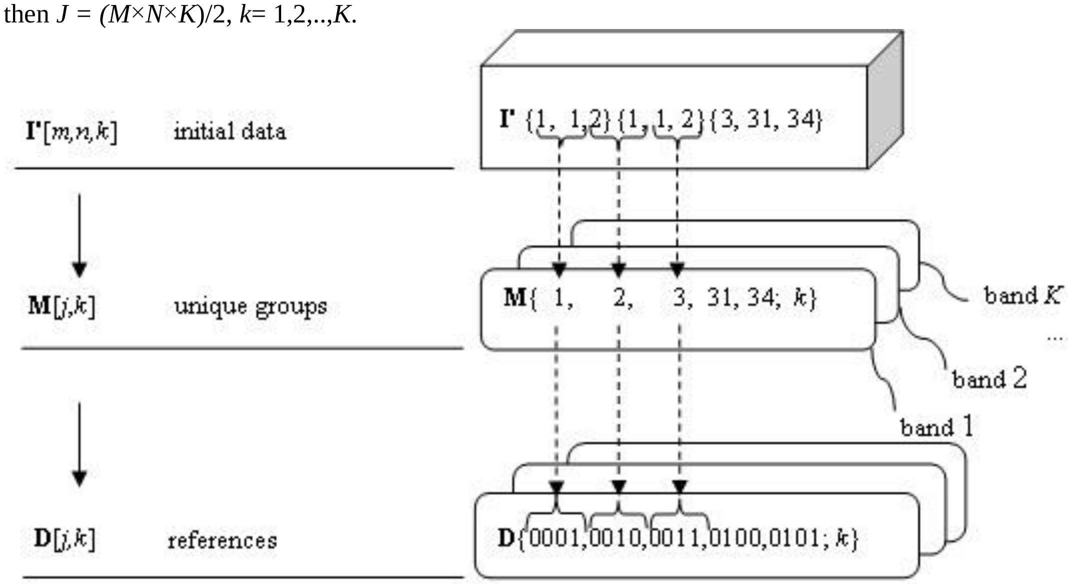

Figure 1. Procedure of forming auxiliary structure of data.

At the end of the second stage, context modeling was used with the known arithmetic coding for compressing data of the file **D**[*j*,*k*] to the archival software **D'**[*j*,*k*].

In order to form an initial hyperspectral AI **I**[*m*,*n*,*k*] from **D'**[*j*,*k*] it is necessary to make a number of transformations opposed to the above-mentioned:

- To make arithmetic decoding of the file **D'**[*j*,*k*] restoring the file **D**[*j*,*k*];
- To find in the file **D**[*j*,*k*] the corresponding references to unique pair groups from the formed structure of the data **M**[*j*,*k*];
- To restore the file **I'**[*m*,*n*,k] containing a file of dependencies **L**[*k*] having counted the absolute values of the file **I'**[*m*,*n*,k] and having restored the initial image **I**[*m*,*n*,k].

III. EXPERIMENTAL RESEARCH

In order to access the effectiveness of the proposed algorithm in what concerns both the point compression ratio as the limits of its application, a number of experiments was perfomed using hyperspectral AI of the system RS *AVIRIS* (table 1) in the format of data of raster geoinformation system *Idrisi Kilimanjaro*. The system *AVIRIS* (*Airborne Visible/Infrared Imaging Spectrometer*) provides 224 spectral images with the wavelength of the band from 400 nanometers to 2500 nanometers. Also, the comparison of the proposed algorithm with the results of experiments received for universal archivers compression algorithms *WinRar*, *WinZip* and compressor *Lossles JPEG* which applies the resolution of compression standard *JPEG* widely used in commercial compression systems were made.

The experiments are made on a computer with the processor *Intel Core i5* 2,5 GigaHerz and RAM 4 Gigabit under operating system *Windows 7* (*updating package 3*).

TABLE I. EXAMPLES OF CHARAC TERISTICS OF TEST DATA (OF THE SYSTEMS RS *AVIRIS*)

Name of hyperspectral AI	**K**	**M × N, pixel**	**File size, byte**
f970619t01p02_r07_sc01	224	100 × 100	6140096
f970619t01p02_r07_sc02	224	200 × 200	36199296
f970619t01p02_r07_sc03	224	300 × 300	81178496
f970619t01p02_r07_sc04	224	400 × 400	144077196
f970619t01p02_r07_sc05	224	500 × 500	210746396
f970619t01p02_r07_sc06	224	624 × 512	281673728

The proposed algorithm was then performed sequence of stages to find the most effective:

Sequence I – Consideration of the correlation (1) $\rightarrow$ forming an auxiliary of data with unique pair groups and references to them (2) $\rightarrow$ arithmetic coding (3).

Sequence II – Forming an auxiliary of data with unique pair groups and references to them (1) $\rightarrow$ arithmetic coding (2).

Sequence III – Consideration the correlation (1) $\rightarrow$ arithmetic coding (2).

Sequence IV – Forming an auxiliary of data with unique pair groups and references to them (1) $\rightarrow$ consideration the correlation (2) $\rightarrow$ arithmetic coding (3).

In order to evaluate the most productive sequence considering the contribution of each of these stages a number of experiments with different variants was conducted (Figure 2).

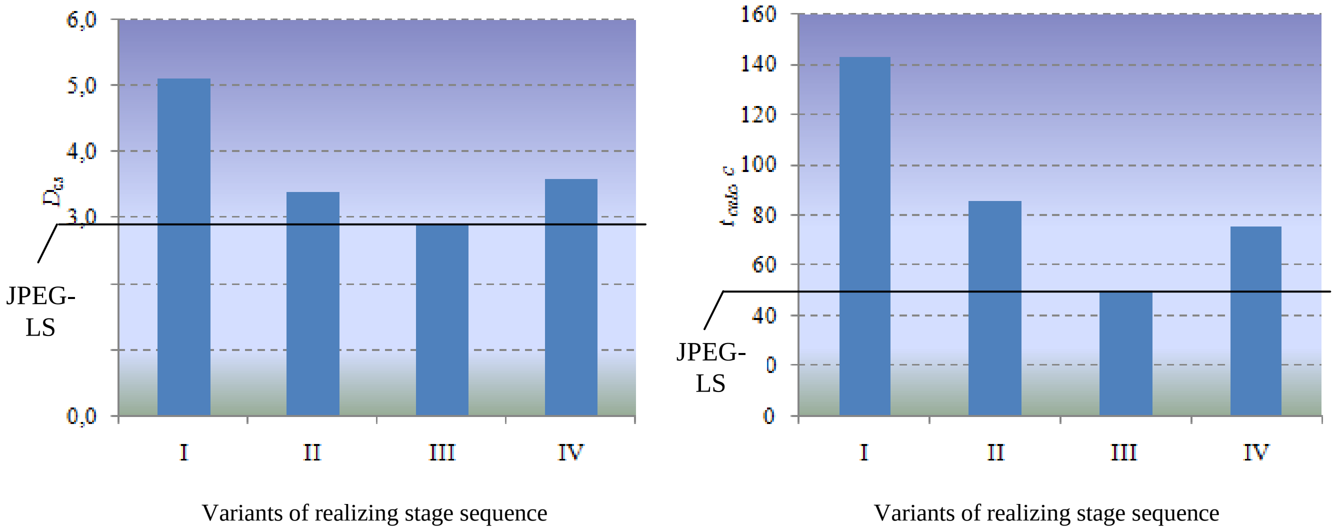

Figure 2. Comparison of indicators of variants of realizing a compression algorithm: a) by exponent of compression D_{cs}; b) by time of calculating work t_{calc}

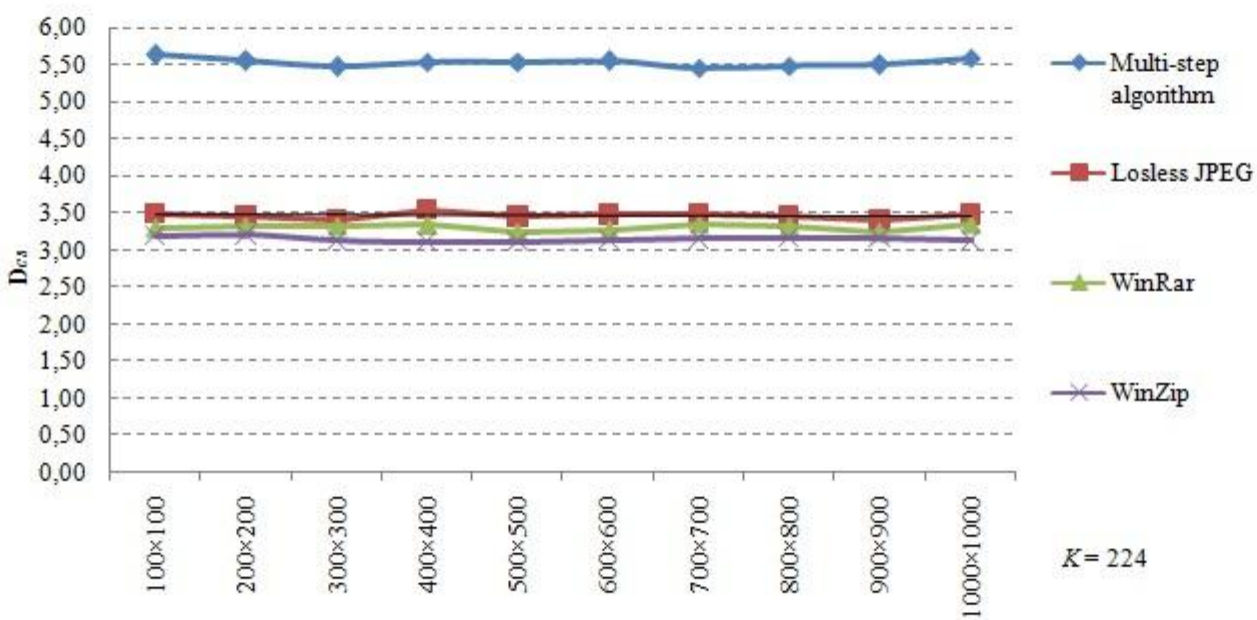

Figure 3. Comparative effectiveness of compression algorithms for different geometrical size of scene.

A fragment of the _results of this experiment is displayed results in the Figure 2, that shows that various stages of the algorithm have different importance while forming a result. In the 1st, 2nd and 4th variants of the stage sequences, the results surpass *Losless JPEG* indifferent degrees (from 25% to 46%).

The best result is achived by sequence I is with the highest exponent of compression. Sequence IV was impaired by the execution of stage it is not expedient to realize the stage (of considering the correlation after the stage of forming auxiliary structures of data) as it leads to increasing a range of data change while counting the difference. The results of realizing the second stage show that absence of the stage of intra-bands correlation leads to an insignificant exponent of compression in comparison with *Losless JPEG*.

Realization of the stage sequence in accordance with the sequence III leads to the most insignificant result as there is no powerful formation of unique pair groups and references to them with creation of corresponding auxiliary structures of data.

In Figure 3, the results of comparative experiments demonstrate the superiority of the proposed algorithm over analogues in exponent of compression D_{cs} at the varied geometrical sizes of hyperspectral AI. At increasing the geometrical size of the scenes, all the investigated algorithms show a steady result which exhibits little to no change.

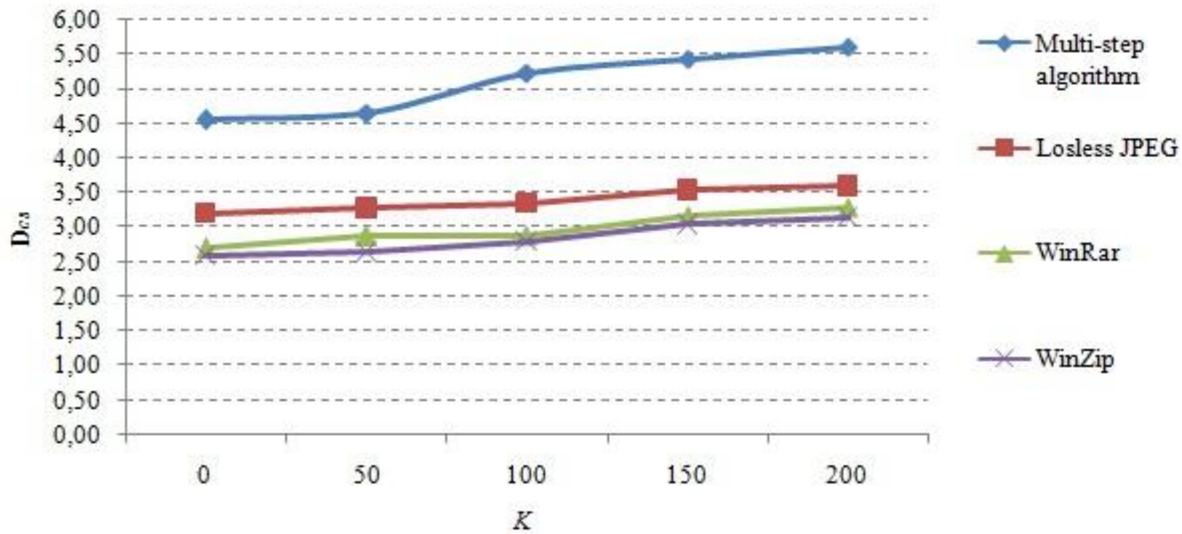

Figure 4. Dependence of compression exponent D_{cs} of algorithms on the number of bands K.

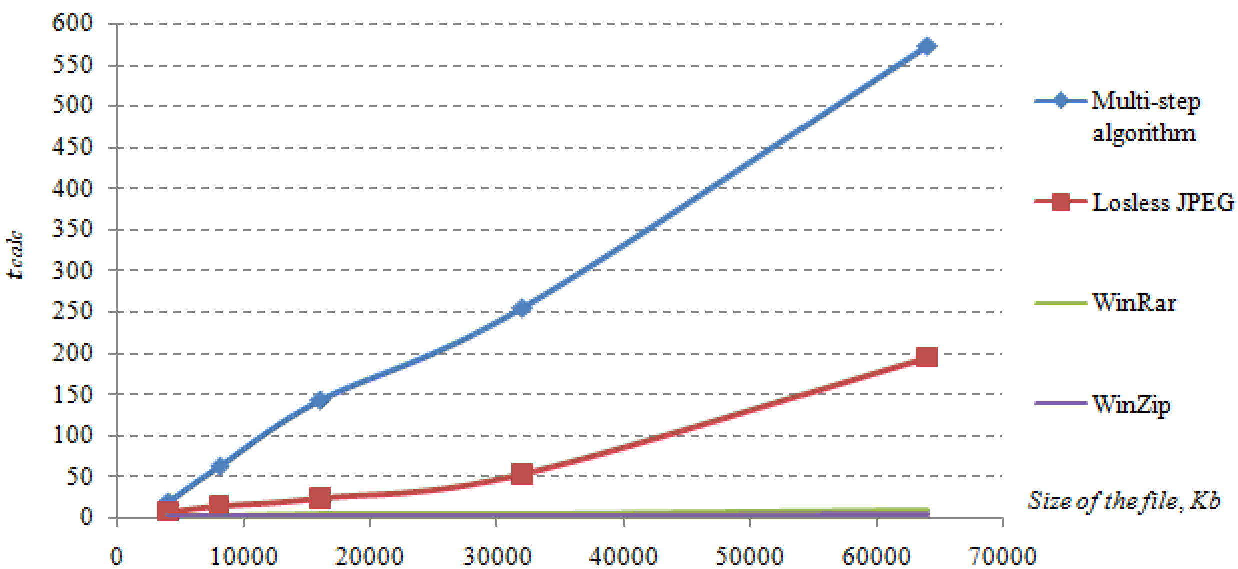

Figure 5. Computational performance of compression algorithms.

Research was conducted to explore the dependence of compression exponent D_{cs} on the number of bands of AI K (Figure 4). Results show that compression exponent D_{cs} is increasing proportionally to the number of bands Kas the redundancy of the data of hyperspectral AI raises.

In conducting comparative research of compression exponents it is necessary to pay attention to calculating expenses of compression algorithms (Figure 5).

As seen in Figure 5, in the proposed algorithm, the calculated effectiveness in comparison with analogues increased 3 fold. This is explained by an improved multi-stage algorithm which is provided to form auxiliary structures of data on AI considering the correlation and the following arithmetic coding. Universal archival software does not take into account the specificity of the data being compressed and does not account for such operations.

IV. Conclusion

1. A multi-stage algorithm for compressing hyperspectral AI was developed. This algorithm considers intra-bands correlation and the preliminary byte data processing, allowing up to 46 % increase in data compression when compared with other algorithms.

2. The analysis of the importance of stages has shown that the stage of preliminary byte processing with formation of auxiliary structures of data allows to improve the result considerably – up to 45 %. The stage of considering intra-bands correlation is less significant. However it allows to lower a range of varied values for operating by smaller spacing, allowing to increase the compression ratio considerably – up to 26 %.

3. The analysis of computing efficiency has shown, that in order to achieve significant results of compression in applying a multi-stage algorithm, high computing expenses are required, conceding to the nearest analogue *Lossless JPEG* up to 3 folds.

This work has been supported by Russian Foundation for Basic Research (grant *№ 14-07-00027a*).

References

[1] V.G. Bondur, “Modern approaches to processing of hyperspectral space images”,Research Institute of space monitoring “Aerospace”, Moscow, 2013, p. 4.

[2] M.A. Popov and S.A. Stankevich, “Optimization methods of spectral bands number in tasks of processing and data analysis of distance remote sensing of the earth”, Scientific center of aerospace earth research, 2003, no. 1, pp. 106-112.

[3] H. Wang, S. D. Babacan, and K. Sayood, “Lossless hyperspectral-image compression using context-based conditional average”, vol. 45, no. 12, 2007, pp. 4187–4193

[4] H. Chengfu, R. Zhang, and P. Tianxiang, “Lossless Compression of Hyperspectral Images Based on Searching Optimal Multibands for Prediction”, vol. 6, no. 2, 2009, pp. 339–343.

[5] Y. Liang, L. Jianping, and Ke G., “Lossless compression of hyperspectral images using hybrid context prediction”, vol.20, no. 7, 2012,. pp. 199–206.

[6] B. Aiazzi, L. Alparone, S. Baronti, C. Lastri, and M. Selva, “Spectral Distortion in Lossy Compression of Hyperspectral Data”, Journal of Electrical and Computer Engineering, no. 10, 2012, p. 8., doi:10. 1155/ 2012/ 850637

[7] L. Cheng-chen and H. Yin-tsung, “Lossless Compression of Hyperspectral Images Using Adaptive Prediction and Backward Search Schemes”, Journal of Information Science and Engineering, no. 27, 2011, pp. 419–435.

[8] B. Aiazzi, L. Alparone, and S. Baronti, “Near-lossless image compression by relaxation-labeled prediction, Signal Process”, no. 11, 2002, pp. 1619–1631.

[9] E. Magli, G. Olmo, and E. Quacchio, “Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC”, IEEE Geoscience and remote sensing letters, no. 1, 2004, pp. 21–25.

[10] B. Aiazzi, S. Baronti, and L. Alparone, “Lossless compression of hyperspectral images using multiband lookup tables”, IEEE Signal Process. Letters, vol.6, no. 16, 2009, pp. 481–484.

[11] B. Penna, T. Tillo, E. Magli, and G. Olmo, “Transform Coding Techniques for Lossy Hyperspectral Data Compression”, IEEE Geoscience and remote sensing letters, vol. 45, no. 5, 2007, pp. 1408–1420.

[12] RarLab. WinRar software system for compress files [http://www.win-rar.com/rarproducts.html]. Retrieved: 15.01.2014.

[13] WinZip. Program of compression for Windows [http://www.winzip.com/ru/prodpagewz.htm]. Retrieved:. 19.01.2014

[14] X. Tang, W. Pearlman, and J. Modestino, “Hyperspectral image compression using three-dimensional wavelet coding”, Proc. SPIE IS&T, 2003, pp. 1037–1047.

[15] B. Penna, T. Tillo, E. Magli, and. G. Olmo, “Progressive 3-D coding of hyperspectral images based on JPEG 2000”, IEEE Geoscience and remote sensing letters. vol. 1, no. 3, 2006, pp. 125–129.

[16] J. Zhang and G. Liu, “An efficient reordering prediction-based lossless compression algorithm for hyperspectral images”, IEEE Geoscience and remote sensing letters.–. vol. 2, no. 4, 2007, pp. 283–287.

[17] J. S. Mielikainen, A. Kaarna, and P. Toivanen “Lossless hyperspectral image compression via linear prediction”, Proc. SPIE 4725, no. 8, 2002, pp. 600–608.

[18] F. Rizzo, B. Carpentieri, G. Motta, and J. A. Storer, “Low-complexity lossless compression of hyperspectral imagery via linear prediction”, IEEE Signal Process. Lett, vol. 2, no. 12, 2005, pp. 138–141.

[19] ISO/IEC 15444-1. JPEG2000 Image Coding System [http:/www.jpeg.org/public/15444-1annexi.pdf]. Retrieved: 10.01.2014.

[20] A. Kiely, M. Klimesh, H. Xie, and N. Aranki, “ICER–3D: A Progressive Wavelet-Based Compressor for Hyperspectral Images”, 2006, pp. 142–164.

[21] L. Gueguen, M. Trocan, B. Pesquet-Popescu, A. Giros, and M.A. Datcu, “Comparison of multispectral satellite sequence compression approaches”, Signals, Circuits and Systems, no. 1, 2005, pp. 87–90.

[22] A.V. Zamyatin and T. D. Chung, “Compression of multispectral space images using wavelet transform and intra-bands correlation”, Journal of Tomsk Polytechnic University, vol. 313, no. 5., 2008, pp. 20–24.

[23] Interregional public organization promoting the market development of geographic information technologies and services “GIS-Association”, [http://www.gisa.ru/1489.html], Retrieved: 15.01.2014.

[24] Arc View GIS [http://gisa.ru/3577.html]. Retrieved: 10.01.2014.

[25] Geo-information systems [http://loi.sscc.ru/gis/default.aspx]. Retrieved: 05.01.14.

[26] D. Vatolin, A. Ratushnyak, M. Smirnov, and V. Yukin, ”Data compression methods”, M.:Dialogue, MIFI, 2003, p. 384.

[27] V.N. Kopylov, “Creation basics of aerospace environment monitoring”. Yekaterinburg: PP “Kontur”, 2006, p. 144.

[28] V.F. Babkin, I.M. Knizhny, and K.E. Chrekin, “Compression of multispectral images with out loss or limited losses’, Modern and perspective developments and technologies in a space instrumentation: Moskva, IKI RAN,. no. 1, 2004, pp. 330–332.

[29] F. Rizzo, B. Carpentieri, G. Motta, and J.A. Storer, “Low-complexity lossless compression of hyperspectral imagery via linear prediction”, IEEE Signal Process. Letters, vol. 2, no. 12, 2005, pp. 138–141.

[30] G. Motta, F. Rizzo, and J.A. Storer, “Hyperspectral Data Compression”, Berlin: Springer, 2006, p. 415.

[31] H. Goto, Y. Hasegawa, and M. Tanaka, “Efficient Scheduling Focusing on the Duality of MPL Representatives,” Proc. IEEE Symp. Computational Intelligence in Scheduling (SCIS 07), IEEE Press, 2007, pp. 57–64, doi:10.1109/SCIS.2007.357670.

[32] D. Salomon and G. Motta, Hand book of data compression, 5th ed., Springer London Dordrecht Heidelberg New York, 2010, p. 42, doi:10. 1007/10.1007/978-1-84882-903-9

[33] A. Robert, “Remote Sensing Models and Methods for Image processing”, Department of Electrical and computer Engeneering College of optical Sciences, University of Orizona, USA, 2007, p. 592.

[34] L. Chang and C. M. Cheng, and Y.L. Chang, “Fast Hyperspectral Image Classification Using Binary Quaternion-Moment-Preserving Thresholding Technique”, IEEE International Geoscience and Remote Sensing Symposium, 2008, Boston, U.S.A., p.2.