Multi-channel data storage format definition for visualization tasks on the example of SPOT-4 images

N Yu Sevastianova[1] and N S Vinogradova[1]

[1]Ural Federal University, Mira street, 19, Ekaterinburg, Russia, 620002

e-mail: n.u.sevastianova@mail.ru

Abstract. One of the features of a remote sensing data storage is the widespread utilization of large-capacity disk arrays. Emergency situations arising from the use of arrays can lead to the fact that the remote sensing data, usually stored in uncompressed form, may become partially damaged. But even with incomplete recovery, this kind of data can be used in the future to solve production problems. However, this recovery is sometimes hampered by incomplete knowledge of the format of the corrupted data. The article describes an approach to automatic recognition of multichannel data interleaving type (BIP, BIL or BSQ) and its application to a recovery of SPOT-4 remote sensing data stored in the segment format "SEG", which were damaged after a disk array failure.

1. Introduction

SPOT is a commercial high-resolution optical satellite earth observation system, which operated in 1998-2013 and aimed at solving environmental and agricultural problems, as well as used in such fundamental areas, as climatology and oceanography [1,2,3,4]. The SPOT-4 device was a multispectral observation system that carried out the survey in four spectral ranges (listed in table 1).

Table 1. SPOT-4 spectral bands [1].

Imaging Band	Spatial Resolution	Wavelength
panchromatic	10 m	0.61–0.68 µm
B1: green	20 m	0.50–0.59 µm
B2: red	20 m	0.61–0.68 µm
B3: near-IR	20 m	0.78–0.89 µm

In 2011-2012, the Space monitoring center of the Ural Federal University received a signal from the SPOT-4 spacecraft and accumulated more than 1 TB of survey data in the original "raw" format (RAW) and the post-processed «SEG» format. After the accident of disk storage in 2016, the data were temporarily lost, but preliminary analysis of the results of the repair work showed that only a part of the data that was stored on failed devices (3 hard disks out of 20) was finally lost. At the same time, the storage file system remained weakly damaged, which made it possible to recover most of the files (including SEG), and the chains of lost fragments have preserved their localization within the file, but were written as zero bytes. Since SPOT-4 data is currently actively used in the educational process, the

question of the visualization possibility and further use of this data turned out to be extremely relevant. In the course of solving this problem, among others, a problem of identifying the method of storing multichannel data have arisen. Different recovering methods for corrupted data are developed last time propose the approaches for recovering different remote sensing data types [5,6], i.e. hyperspectral data [7] and SAR data [8], but the peculiarity of a problem considered here is that data are multispectral and they have the specific corruption type (randomly located horizontal stripes of zeros pixels). There are three main ways to record data in multichannel images [9]:
- BIL (byte interleaving by line): channels separated by lines;
- BIP (byte interleaving by pixel): pixel-by-pixel recording of spectral channels;
- BSQ (band sequential): per-channel recording.

The presented storage methods have an associated header, which is usually represented in ASCII encoding. This header contains basic information about the image, such as the format, the number of rows and columns, the presence of a color map, can also contain auxiliary information, such as geographic reference, authorship, imaging time.

BIL uses a method of recording multi-channel images in which rows of values for each channel are written sequentially. For example, in the case of a three-channel image, all three data channels are first written to the first row, then to the second row, and so on until the last row of the image is written.

BIP method implies that the data for each pixel is written sequentially. For example, when you record a three-channel image, the data for the first, second, and third channels is recorded for the first pixel of the first column; then the channel data is written for the first pixel of the second column, and so on.

For per-channel recording (BSQ), the information for each channel is recorded sequentially. In other words, the data of all pixels of the first channel is recorded first, then all pixels of the second channel are recorded and so on [9]. Automatic recognition of correct data interleaving type makes it possible to organize the on-flow damaged data recovery.

2. The algorithm for the SPOT-4 data analysis

As reference information, the sizes of images in bytes (which can be completely different) and the number of SPOT-4 channels equal to four are used. For data analysis in MATLAB, an algorithm is developed that allows you to select and analyze the method of multispectral data storage. At the first stage, factorization of file size into the prime factors was performed, which allowed splitting the image size to determine the number of columns and rows. It showed, that line size for SEG data is a multiple of 3057, where the first 57 bytes of each channel contains metadata. Another approach may imply the one-dimensional Fourier transform and spectral peaks analysis. The peaks positions correspond to the image number of columns, and the latter makes it possible to determine the image size, but the determination of interleaving type may be inconveniently.

In order to set the desired data storage method, the images are opened with all three possible options. Visual comparison of opened images revealed significant spatial differences (figure 1).

The image that is opened in the assumption that the data stored as BIL, contained stripes along the image columns, which deal with the violation of pixels interleaving method. The image, opened under the assumption of BSQ way, also carries an incorrect pixel striping, that can be observed in the figure 1c (more stripes along image diagonals). An image opened under the assumption that the data is stored in a BIP manner represents the correct striping (figure 1a). Thus, for SPOT-4 remote sensing data, the BIP is the correct way to store multispectral data.

2.1. Analysis of the characteristics of the BIL format

As noted earlier from the visual mapping, an image opened under the assumption using the BIL storage method, has expressed bars in the direction of the image columns. In order to emphasize the bands, spatial filtering methods were used. The following types of spatial filters have been tested [10,11,12]:
- Laplace filters;
- line detection filters;
- Sobel operator;

• Roberts cross filter.
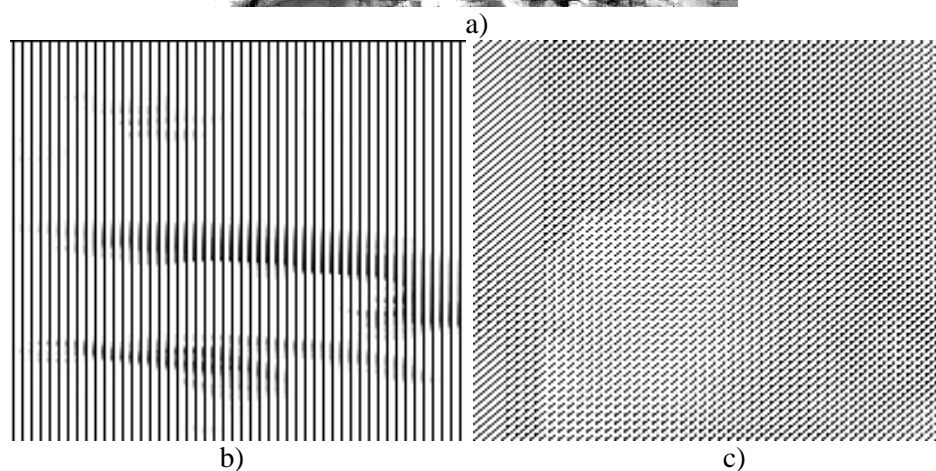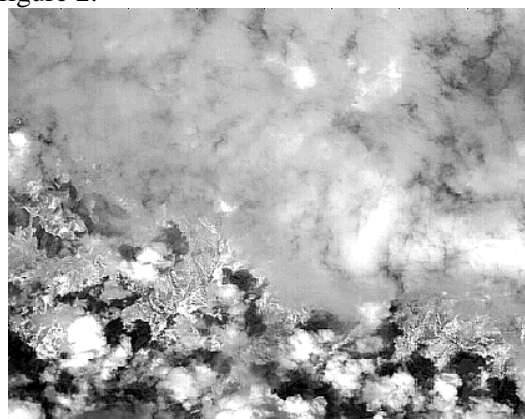The results are shown in figure 2.

a)

b)   c)

Figure 1. Image fragment: a – opened as BIP; b – opened as BIL; c– opened as BSQ.

Variance is calculated for all filtering results (for different channels and different fragments). The variance along the row characterizes the spread of image brightness, thus being a quantitative indicator of the presence of bands oriented along the columns of the image that occurs when the image is opened under the assumption that the storage method is BIL. The maximum variance value is obtained by using the Sobel filter. Figure 3 presents the results of variance calculation for images that are open in the assumption of BIP and BSQ, with the application of the Sobel operator. The maximum variance of the image opened in assumption correct method of storage, do not exceed any fragments the maximum variance of another image opened in the assumption is a wrong way to storage. As can be seen, the image variance, opened with the correct pixel alternation method, does not have sharp changes in comparison with the open with the help of BIL and seems to be smoother. Thus, on the basis of visual comparison and analysis of variance, one can conclude that the Sobel operator is the best filter in the problem of line selection in the direction of image columns.

Figure 4 shows fragments of filtered images using the Sobel operator. The image opened in the assumption method of BIP storing, has no sharp boundary changes, while the image that is open in the BIL assumption, shows sharp stripes in the column direction, which is expressed quantitatively in the calculation of the variance of the image. Thus, this feature can be used as an indicator of identification of the storage format correctness for SPOT-4 data and, possibly, other remote sensing data types.

2.2. Analysis of the BSQ format characteristics
As noted earlier, the image, which is opened under the assumption of the BSQ storage method, an arbitrary pixel alternation is expressed, which looks like a strip along the diagonals. In order to emphasize the stripes, spatial filtering methods are also used. In the course of the work the various filters presented earlier were tested, as well as others, such as gradient filters and filtration using

homogeneity filter (Figure 5). Since the Sobel Operator and the Robinson cross operator are only used to highlight horizontal or vertical lines, the image was rotated 45 degrees after filtering. The angle four-channel data is 45 degrees.
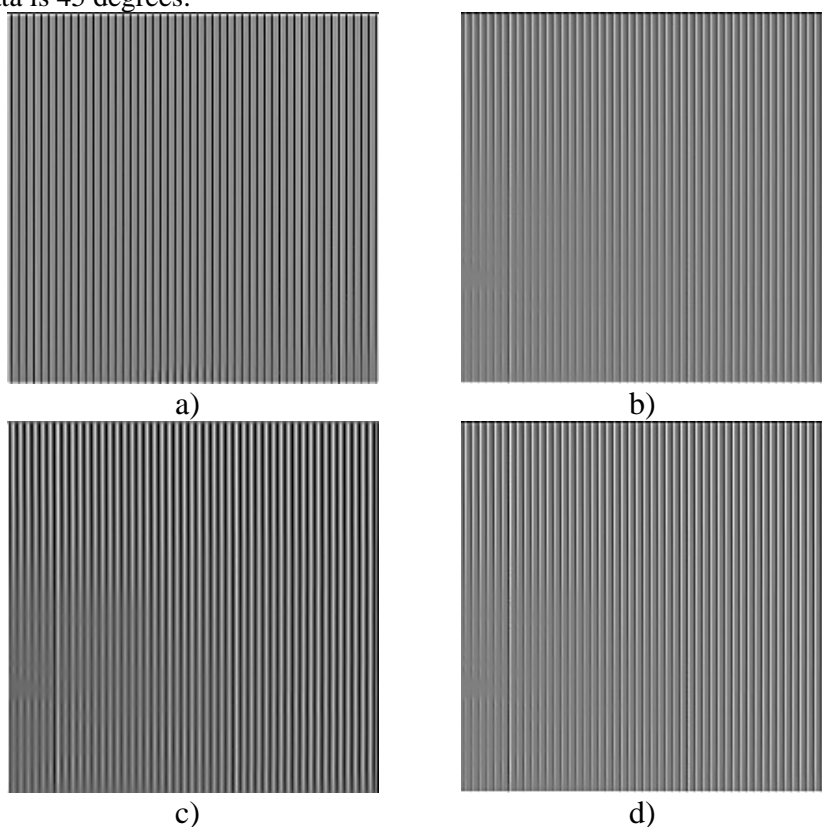
a)

b)

c)

d)

Figure 2. Image filtering result (BIL): a) Laplace filter; b) line detection filter; c) Sobel filter; d) Roberts cross operator.
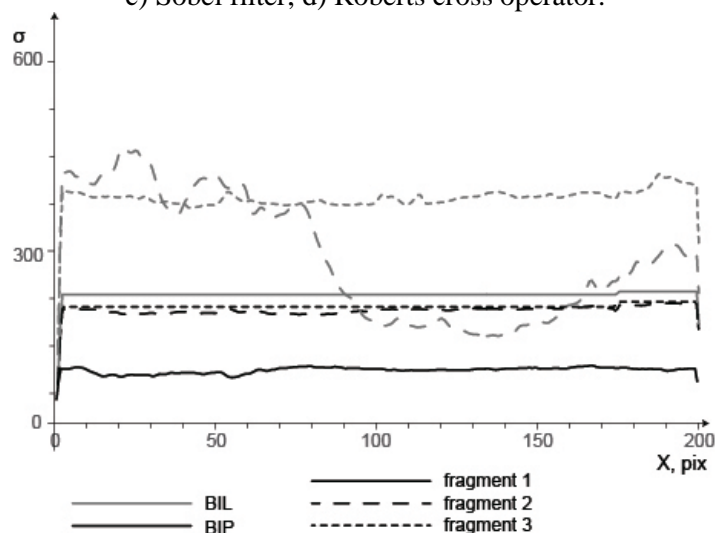
Figure 3. Variance calculations for different fragments opened as BIP and BIL.

For all filtering results, the variance is also calculated, which shows that the Sobel operator provides maximum contrast between the images opened under the assumption of BIP and BSQ. Figure 7 presents the results of variance calculation for images that are opened in the assumption BIP and BSQ, with the help of the Sobel operator. The variance of a data fragment opened by the BIP method is smoother than the variance of another method. Thus, on the basis of visual comparison and analysis of variance, it can be concluded that the Sobel operator in this case also emphasizes the selection of

lines in the direction of the image columns. Figure 6 shows the image fragments filtered by the Sobel operator. The image opened in the assumption with the BSQ method, has vertical stripes, and for the other image that is opened in the assumption with BIP, this is not observed.
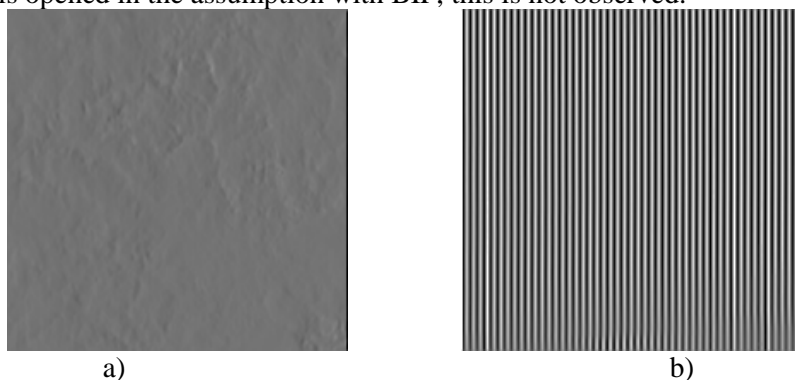

a)   b)

Figure 4. Application of Sobel operator: a – image opened as BIP; b – image opened as BIL.

The example of SPOT-4 recovered data is shown in figure 8. The horizontal black stripes correspond to the data segments, which were located at the damaged disk drives. Vertical stripe at the left edge has the width about 57 bytes (per channel) and contains service information and metadata. The relative number of damaged stripes is about 13.8%, and it fits the number of damaged disks (3 of 20).

3. Results

An algorithm was developed and implemented in MATLAB, that allows one to determine automatically the correct way to multispectral data storage method in terms of pixel interleaving. The algorithm was tested on the example of SPOT-4 remote sensing data stored in post-processed SEG-format. The algorithm is based on Sobel spatial filtering, which emphasizes the geometric features of the image opened in the correct way. In case of an incorrect way, the variance calculated along the line of the image gives inflated values, thus it becomes possible to identify an adequate way of data interleaving type. The algorithm shows reliable results on eighty-seven different fragments of SPOT-4 remote sensing data in four different channels. The proposed algorithm was used for SPOT-4 data restoration after the disk storage failure.
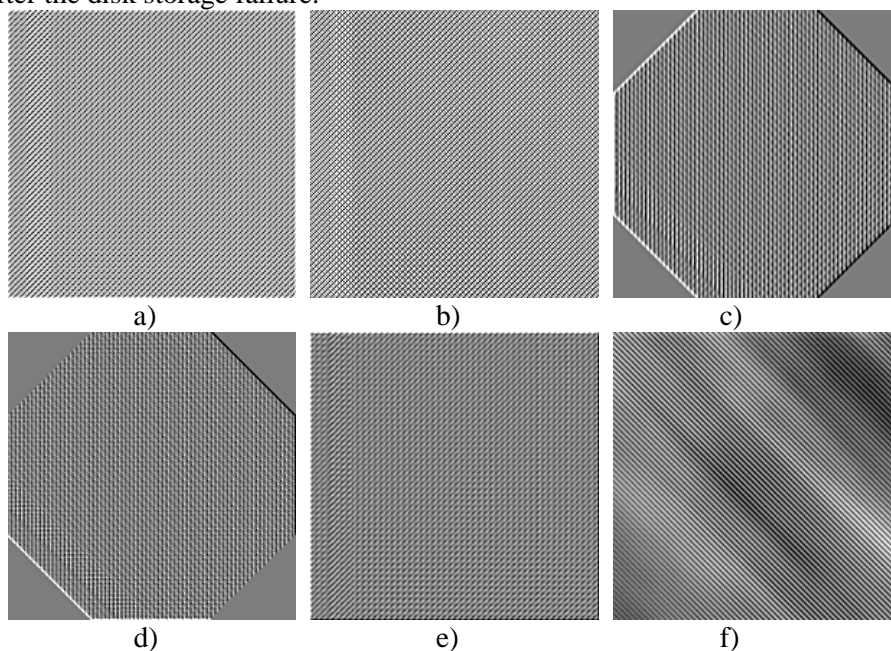

a)   b)   c)

d)   e)   f)

Figure 5. Scene filtration results: a) Laplace filter; b) lines detection filter; c) Sobel operator; d) Roberts cross operator; e) gradient filter; e) homogeneity operator.
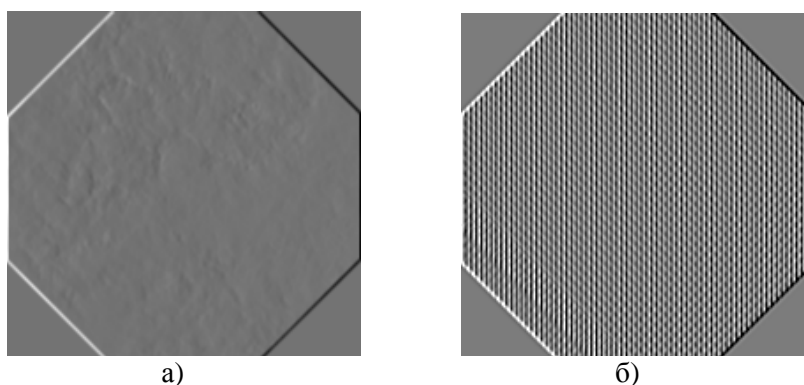
Figure 6. Application of Sobel operator: a) for scene opened as BIP; b) for scene opened as BSQ.
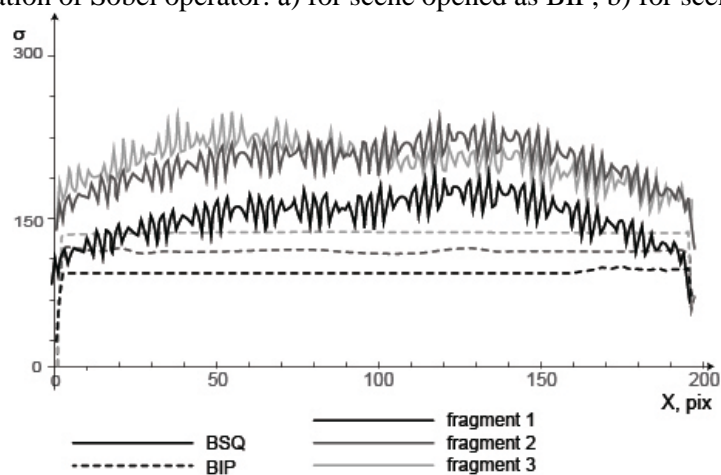

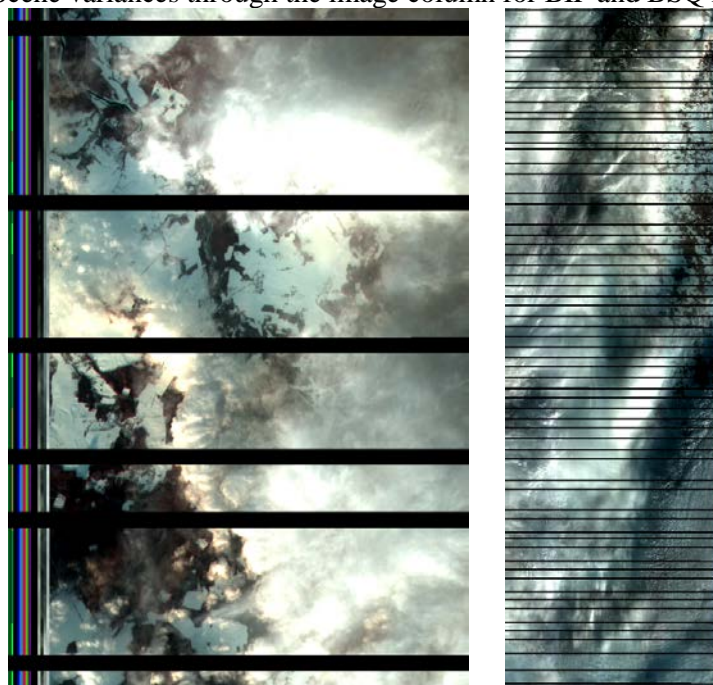**Figure 7.** Scene variances through the image column for BIP and BSQ interleaving.

Figure 8. SPOT-4 recovering: a) image fragment; b) full scene.

4. References

[1]   Arnaud M, Leroy M 1991 *ISPRS Journal of Photogrammetry and Remote Sensing* **46(4)** 205-
       215

[2] Yurikova E A, Kokutenko A A and Sukhinin A I 2008 Research of opportunities of application of data SPOT-4 for interpretation the areas of vegetation damaged by fires *Vestnik Sibirskogo gosudarstvennogo aerokosmicheskogo universiteta im. akademika M.F. Reshetneva* **4** 75-78

[3] Plotnikov D Y, Kolbudaev P A and Bartalev S A 2018 Identification of dynamically homogeneous areas with time series segmentation of remote sensing data *Computer Optics* **42(3)** 447-456 DOI: 10.18287/2412-6179-2018-42-3-447-456

[4] Boori M S, Paringer R, Choudhary K and Kupriyanov A 2018 Comparison of hyperspectral and multi-spectral imagery to building a spectral library and land cover classification performance *Computer Optics* **42(6)** 1035-1045 DOI: 10.18287/2412-6179-2018-42-6-1035-1045

[5] Dong J, Yin R, Sun X, Li Q, Yang Y and Qin X 2019 Inpainting of Remote Sensing SST Images with Deep Convolutional Generative Adversarial Network *IEEE Geoscience and Remote Sensing Letters* **16(2)** 173-177

[6] Chang Y, Yan L, Wu T and Zhong S 2016 Remote Sensing Image Stripe Noise Removal: From Image Decomposition Perspective *IEEE Transactions on Geoscience and Remote Sensing* **54(12)** 7018-7031

[7] Eason D, Andrews M 2014 Compressed hyperspectral image recovery via total variation regularization assuming linear mixing *IEEE International Conference on Image Processing* 620-624

[8] Biondi F 2017 Recovery of Partially Corrupted SAR Images by Super-Resolution Based on Spectrum Extrapolation *IEEE Geoscience and Remote Sensing Letters* **14(2)** 139-143

[9] Bossler J D 2010 *Manual on geospatial science and technology. Second edition* (Boca Raton: CRC Press) p 832

[10] Sirota A A, Solomatin A I 2008 Statistical algorithms for detecting the boundaries of objects in the image *Vestnik VGU* **1** 58-64

[11] Gruzman I S 2012 Synthesis of optimal differentiators for the locally oriented texture detection algorithm *Computer Optics* **36(1)** 109-115

[12] Zhuravel Y N, Fedoseev A A 2013 The features of hyperspectral remote sensing data processing under environment monitoring tasks solution *Computer Optics* **37(4)** 471-476