Optimization of the Composition of Standards in Recognition and Compression Algorithms of Hyperspectral Images[*]

Leonid Lebedev

Lobachevsky state university of Nizhny Novgorod,
Nizhny Novgorod, Russia
lebedev@pmk.unn.ru

Abstract. The paper proposes a solution to the problem of minimizing the number of standards in order to increase both the compression coefficient of hyperspectral images (HSI) and the speed of correlation extreme compression methods (CEM). As modifications of the CEM, randomized and differential compression algorithms are offered. The randomized and difference algorithms are based on the hypothesis of spatial compactness of pixels located in local regions of the image matrix. This means that when a new template is formed based on an unrecognized pixel, there is a high probability of using a pixel that lies near the boundaries of the coverage areas of the existing templates, which leads to their increase. In order to reduce the influence of spatial compactness of pixels on the formation of standards, a methodology based on changing the sequence of recognized pixels is proposed. In a randomized algorithm, a row of the matrix is randomly determined for this, on the basis of which a sequence of recognized pixels is generated by a random column generator. In the difference algorithm of compression, the row number of the matrix is determined by the rule for finding the members of an arithmetic progression with a given difference. For the selected line a sequence of recognizable pixels is formed on the same principle. It should be noted that line-by-line pixel recognition in the self-learning mode allows compressing HSI of almost any volume. The effectiveness of the created algorithms is demonstrated on two fragments of real HSI. A comparative analysis of all three compression algorithms in terms of the quantitative composition of the obtained standards is presented.

Keywords: Hyperspectral Image, Correlation Extreme Methods, Similarity Estimates, Recognition, Compression, Random Pixel Samples.

---

1   Introduction

In [1], the basics of correlation extreme methods of recognition and compression in self-learning modes that are invariant to a given type of transformation, when hyperspectral images act as the source information. The HSI is represented as a three-dimensional cube, each pixel of which is described by the response values in the corresponding spectral channels. Representing HSI pixels as vectors (points) $\mathbf{y} = (y_1, y_2, ..., y_n)$ in a multidimensional linear space $\mathbb{R}^n$, where n it corresponds to the number of channels of the spectrometer, allows us to obtain similarity estimates for recognition and compression methods that are invariant, respectively, to the identity transformation ε_m^I and $\hat{\varepsilon}_m^I$; to the similarity transformation ε_m^S and $\hat{\varepsilon}_m^S$; to the offset ε_m^D and $\hat{\varepsilon}_m^D$; to the scaling and offset ε_m^T and $\hat{\varepsilon}_m^T$. The difference in similarity estimates for recognition and compression methods is that the conversion operators adduce the source pixel $\mathbf{y}$ to the reference pixel $\mathbf{y^e} = (y_1^e, y_2^e, ..., y_n^e)$ during recognition, while compression, on the contrary, returns the restored value $\mathbf{y}$ based on the corresponding conversion of the reference pixel $\mathbf{y^e}$. Note that for any pair of pixels $\mathbf{y}$ and $\mathbf{y^e}$ for identical conversion and offset, the similarity estimates for the recognition and compression methods are the same. To increase the compression ratio of the HSI, two directions can be easily identified: first, with the choice of compression method, and second, with a decrease in the number of standards. When compressing based on recognition with self-learning, reducing the number of standards also increases the speed of reduction algorithms. In [1], to reduce the number of standards, an algorithm based on the idea of solid stacking by means of coverage zones of standards of the entire set of pixels is proposed. For the identical transformation, the formation of a new standard based on an unrecognized pixel is carried out in accordance with the formula (1)

$$y_v^e = y + \sum_{i \in Q} \vec{l}_i, \quad \vec{l}_i = (y - y_i^e) \cdot (\lambda \cdot \sqrt{\frac{\delta}{\hat{\varepsilon}_{m,i}}} - 1), \tag{1}$$

where Q is the set of standards for which the pixel lies in the region of their solid stacking; δ is the compression threshold; and λ is a parameter whose value depends on the space dimension n and is responsible for the implementation of the regions of solid stacking.

   In a space of dimension $n = 2$, the formation of a new standard $\mathbf{y_3^e}$, the location of which ensures optimal solid stacking in accordance with formula (1), is illustrated in Fig. 1. The use of this method for optimizing the number of standards in compression algorithms allows reducing their number by magnitude up to 5%. However, there are difficulties in its practical application. The difficulty of applying this method lies in the choice of the value of the parameter h for the implementation of the optimal continuous stacking. Obviously, a violation of the principle of solid stacking or excessive intersection of the coverage areas of the standards leads to a decrease in the efficiency of the optimization method. In turn, the analysis of the formation of standards shows that

when using a sequential enumeration of pixels (television scan), the intersection of the coverage areas of the standards is a significant factor in their excess replenishment.
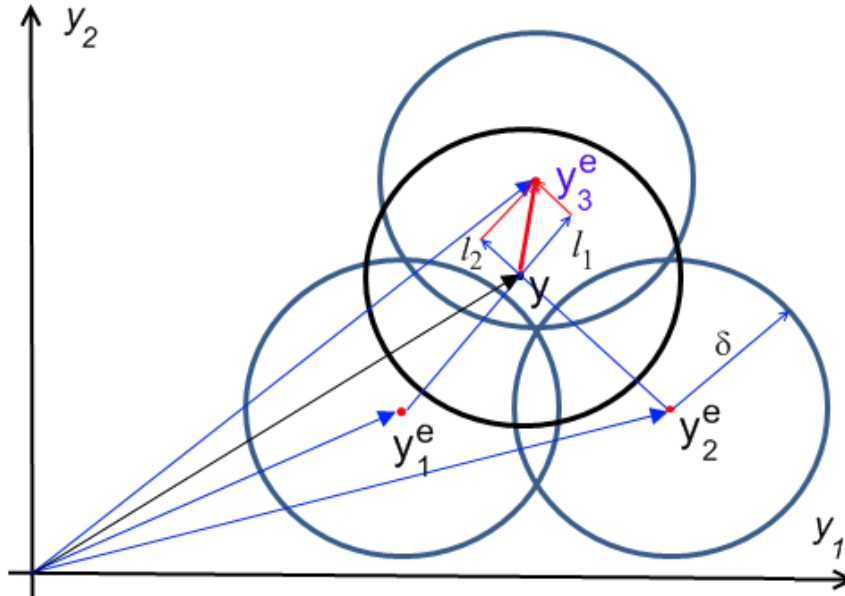

Fig. 1. Illustration of the method of minimizing the number of standards in the compression algorithm with an identical estimate

Therefore, in order to focus on pixels that are more distant in space from the reference signatures, it is proposed to select pixels for compression at random. Thus, the randomness of the choice of pixels can provide continuous coverage with fewer standards. The same idea in solving other problems can be traced in [2-4].

2 Methods of minimizing the number of standards

Consider two algorithms, randomized and difference, that minimize the number of standards relative to the original HSI compression algorithm, implemented on the basis of the correlation-extreme recognition method with self-learning.

2.1 The randomized compression algorithm for hyperspectral images

The main difficulty in implementing the compression algorithm when forming a random pixel recognition sequence is the large amount of HSI, which does not allow to store all the information in many programming systems in RAM. Therefore, the selection of a pixel at random in the proposed randomized compression algorithm of the HSI will be carried out in two stages. First, a row of an image matrix is determined by a random number generator, and then a sequence of pixels of this row is randomly

generated. At each step, the values of random numbers are controlled in order to exclude the possibility of their coincidence. Since at the second step, the selection of pixels is carried out within one row of the matrix, the dimensions of the HSI are not critical for this implementation of the algorithm.

2.2 The difference algorithm of compression for hyperspectral images

An alternative algorithm to the above is the difference algorithm of compression of the HSI. Here, in the proposed algorithm, the row is selected according to the principle of obtaining the members of the arithmetic progression with a given difference d. Line numbers for compression in this algorithm are formed by the formula

$$Ni = (i \cdot d)\%Kr + ent(i \cdot d/Kr), \quad i \in \{0,1,\dots,Kr-1\}, \tag{2}$$

where Kr is the number of rows of the HSI matrix. To determine the sequence pixel numbers in a row, one can use the same formula (2), replacing the value of the parameter Kr in it by the value Ks equal to the number of columns and possibly changing the difference d. Thus, by changing the sequence of lines and pixels in them in accordance with the proposed algorithm, it is possible to reduce the correlation of current pixels. And this implies a decrease in the number of standards for continuous coverage.

3 Experimental research

Experimental studies to minimize the number of standards in the proposed algorithms, randomized and differential algorithm, were carried out on two fragments of HSI with a significantly different composition of objects.

3.1 Experiments on the HSI fragment f100520t01p00r12

The first fragment was represented by lines 251 through 550 of the f100520t01p00r12 HSI file of the AVIRIS spectrometer based on 224 frequencies. The number of columns of the fragment matrix was 813 pixels. The spatial resolution was 17.3 m. The image of this fragment is shown in Fig. 2. For all four types of similarity ratings, this fragment was compressed under various conditions. The original compression algorithm was tested on a fragment when specifying various lines that were taken as start lines. Basically, the step of changing the start lines was 10 units. The way to select the current lines further after the start was sequential. As a result of the experiments, for each assessment, a set of Ms was obtained, the elements of which are the values of the number of standards formed during compression of the fragment. The boundary values, the average number of standards, and the variance were found for this set.

Fig. 2. Colored image of the f100520t01p00r12 HSI fragment

For a randomized algorithm, similar characteristics were obtained. For the difference algorithm, the number of generated standards was obtained with the difference value $d = 10$. The results obtained are summarized in Table 1.

Table 1. The results of experiments on the f100520t01p00r12 HSI fragment

Adaptive compression algorithm, $\sigma = 2\%$		with similarity estimate $\hat{\varepsilon}_m^I$	with similarity estimate $\hat{\varepsilon}_m^S$	with similarity estimate $\hat{\varepsilon}_m^D$	with similarity estimate $\hat{\varepsilon}_m^T$
Initial algorithm	Interval [min, max]	[544, 559]	[538, 546]	[421, 430]	[195, 209]
	The average number of standards	**551.53**	**242.5625**	**426.3125**	**200.1875**
	Standard Deviation	4.13	2.5487	2.1424	3.8278
Randomized algorithm	Interval [min, max]	[521, 552]	[228, 246]	[407, 421]	[191, 204]
	The average number of standards	**536.69**	**236.7333**	**413.8125**	**196.5625**
	Standard Deviation	7.98	4.3660	5.2585	3.3721
Difference algorithm, d=10	The number of standards	**525**	**238**	**409**	**195**

From the data given in the table, it follows that the randomized and difference algorithms are superior to the original compression algorithm in terms of minimizing the number of standards needed to restore a fragment of the HSI with an error not exceeding 2% of the pixel norm. In turn, the difference algorithm is more efficient than the randomized algorithm by the same criterion, although this is mainly for assessing similarities in an identical transformation. However, as follows from the obtained boundaries of the change in the number of standards in the experiments performed, in some cases the randomized algorithm is superior to the difference algorithm, although on average it can be inferior to it.

3.2 Experiments on the HSI fragment MoffettField

The second fragment was formed on the basis of lines 101 to 600 of the MoffettField HSI. The number of matrix columns is 753, and the number of channels used by the AVIRIS spectrometer was 224. This fragment is shown in Fig. 3. The difference of this fragment from the previous one lies in the greater diversity of the underlying surface and, therefore, the need to form a much larger number of standards.

Fig. 3. Image of MoffettField HSI fragments (lines from 101 to 600)

The experiments on this fragment were carried out according to the same scheme. The research results are shown in table 2. As follows from the results in the tables, when compressing the HSI using a similarity estimate that is invariant with respect to the identical transformation $\hat{\varepsilon}_m^I$ or the displacement $\hat{\varepsilon}_m^D$, preference should be given to the difference algorithm. To evaluate the similarity invariant with respect to the similarity transformation $\hat{\varepsilon}_m^S$ or the similarity transformation with offset $\hat{\varepsilon}_m^T$, both compression algorithms showed almost identical results in minimizing the number of standards.

However, when choosing a compression algorithm for HSI in this case, a randomized algorithm should be preferred because of the lack of settings in it. In the difference algorithm, the value of the parameter d should be estimated mainly from the data on the spatial resolution of the spectrometer used, with the aim of choosing the sequence of less correlated pixels.

Table 2. The results of experiments on the MoffettField HSI fragment

Adaptive compression algorithm, $\sigma = 5\%$		with similarity estimate $\hat{\varepsilon}_m^I$	with similarity estimate $\hat{\varepsilon}_m^S$	with similarity estimate $\hat{\varepsilon}_m^D$	with similarity estimate $\hat{\varepsilon}_m^T$
Initial algorithm	Interval [min, max]	[9270, 9551]	[3975, 4037]	[6242, 6312]	[3171,3240]
	The average number of standards	**9362.23**	**4008.10**	**6271.00**	**3199.00**
	Standard Deviation	85.1121	21.6492	24.8193	21.46
Randomized algorithm	Interval [min, max]	[9082, 9246]	[3889, 3983]	[6152, 6226]	[3147, 3178]
	The average number of standards	**9174.28**	**3954.6333**	**6177.500**	**3159.3**
	Standard Deviation	34.1614	23.7982	24.5937	11.63
Difference algorithm, $d=10$	The number of standards	**9184**	**3956**	**6105**	**3149**

4 Conclusion

The hypothesis of a possible reduction in the number of standards as a result of a change in the sequence of pixels during compression of the HSI was confirmed by experiments. As a result, on the basis of the correlation extreme method, two algorithms were created, randomized and differential compression algorithms, which reduce the number of standards by 2-5% and thereby increase the compression coefficient of the HSI, as well as the speed of the procedures.

References

1. Lebedev L.I.: Geometrical aspects of correlation-extreme methods of object recognition and HSI compression. In: Proceedings of the 6th International Conference on Information Technology and Nanotechnology (ITNT-2020), pp. 229-238. Samara National University, Samara (2020)..
2. Svirnov, S.I., Mikhailov, V.V., Ostrikov, V.N.: Application randomized method of principal components for hyperspectral data compression. J. Modern problems of Earth remote sensing from cosmos, 11(2), 9-17 (2014).
3. Borzov, S.M., Guryanov, M.A., Potaturkin, O.I.: Study of the classification efficiency of difficult-to-distinguish vegetation types using hyperspectral data. J. Computer Optics, 43(3), 464-473 (2019).
4. Bibikov, S.A., Kazanskiy, N.L. Fursov, V.A.: Vegetation type recognition in hyperspectral images using a conjugacy indicator. J. Computer Optics, 42(5), 846-854 (2018).