

## **An Optimal Data Loss Compression Technique for Remote Surface Multiwavelength Mapping**

Sergei V. Vasilyev

*Solar-Environmental Research Center, P.O. Box 30, Kharkiv, 310052, Ukraine*

**Abstract.** This paper discusses the application of principal component analysis (PCA) to compress multispectral images taking advantage of spectral correlations between the bands. It also exploits the PCA's high generalizing ability to implement a simple learning algorithm for sequential compression of remote sensing imagery. An example of compressing a ground-based multiwavelength image of a lunar region is given in view of its potential application in space-borne imaging spectroscopy.

### **1. Introduction**

Growing interest in lossy compression techniques for space-borne data acquisition is explained by the rapid progress in sensors, which combine high spatial resolution and fine spectral resolution. As the astronomical community becomes aware of multispectral and hyperspectral imaging opportunities, the airborne land sensing heritage (e.g., Birk & McCord 1994; Rao & Bhargava 1996; Roger & Cavenor 1996) can also valuably contribute to future development of the space related instrumentation and data processing methods. On the other hand the on-board storage capacity and throughput of the space-to-ground link are limited, especially for distant and costly astronomical missions. This constrains direct data transmission and the application of traditional image compression algorithms requiring computationally unreasonable expense and inputs, which are not often available on space observatories. In addition, to meet rational operating of the data transmission channels the raw data need on-board compression prior to their down-linking and an option to compress the data flow bit by bit as it is generated onboard without preliminary image store.

Intuitively, when dealing with the hyperspectral/multispectral imagery, one can expect potential benefits from sharing some information between the spectral bands. As a matter of fact, in the case of high spectral correlations between the bands much of information is redundant, that affords a good opportunity for application of lossless compression.

This paper focuses on the ability of principal component analysis (PCA) to reduce dimensions of a given set of correlated patterns and suppress the data noise through data representation with new independent parameters (Vasilyev 1997). It also implements a simple algorithm, which involves some a priori information about the spectral behavior of the object being studied through

preliminary learning of the principal components for their further application for practically lossless image compression in a computationally inexpensive way.

## 2. Essential Concepts

The starting-point for our approach and for the application of the PCA compression to the multispectral/hyperspectral imagery is a spectral decorrelation transformation at each pixel of the frame on the scanning line across the bands. In other words, let us represent a multispectral image set taken in  $m$  different wavelengths as the assemblage of  $n$  spectra, where  $n$  corresponds to the number of pixels in each single-band picture. Then, let  $A$  be the matrix composed of the signal values  $A_{ij}$  of the  $i$  spectrum in the  $j$  band.

Obviously, all the data, namely each signal value  $A_{ij}$  can be easily restored in the  $m$ -dimensional basis simply by the linear combinations of  $l \leq m$  eigenvectors (principal components) obtained for the covariance matrix  $AA'$ :  $A_{ij} = \sum_{k=1}^l \lambda_{ik} V_{jk}$ , where  $k$  is the principal component number,  $V_{jk}$  is the  $j$ -th element of the  $k$ -th eigenvector and  $\lambda_{ik}$  is the corresponding eigenvalue at the  $i$ -th spectrum.

Generally,  $m$  eigenvectors are needed to reproduce  $A_{ij}$  exactly. Nevertheless, PCA possesses an amazing feature for the eigenvalues sorted in value-descending order: the eigenvectors corresponding to the first, largest eigenvalues bear the most physical information on the data, while the rest account for the noise and can be neglected from further consideration (Genderon & Goddard 1966; Malinowski 1977). Thus, utilizing  $l \ll m$  eigenvectors yields a significant compression of the lossy type and allows the compression rate to be adjusted according to the specific tasks.

Another important feature of PCA that makes this method suitable for the remote sensing applications lies in its powerful generalizing abilities (Liu-Yue Wang & Oja 1993). This feature allows us to describe, using the principal components obtained on the basis of a relatively small calibration data set, a much larger variety of data of the same nature (e.g., Vasilyev 1996; Vasilyev 1997).

## 3. The Data Used

The lunar spectral data obtained by L. Ksanfomaliti (Shkuratov et al. 1996; Ksanfomaliti et al. 1995) with the *Svet* high-resolution mapping spectrometer (Ksanfomaliti 1995), which was intended for the Martian surface investigations from the spacecraft *Mars 94/96*, and the 2-meter telescope of the Pic du Midi Observatory were selected and used for testing the PCA compression. The 12-band spectra (wavelengths from  $0.36\mu m$  to  $0.90\mu m$ ) were recorded in the "scanning line" mode during the Moon rotation at the phase angle of  $45^\circ$ , so that the data acquisition method was similar to the spacecraft scanning. The resulting 19-pixel-wide images were composed of these spectra separately for each band and put alongside in the gray scale (see Figure 1a), where the lower intensity pixels correspond to the higher signal level.

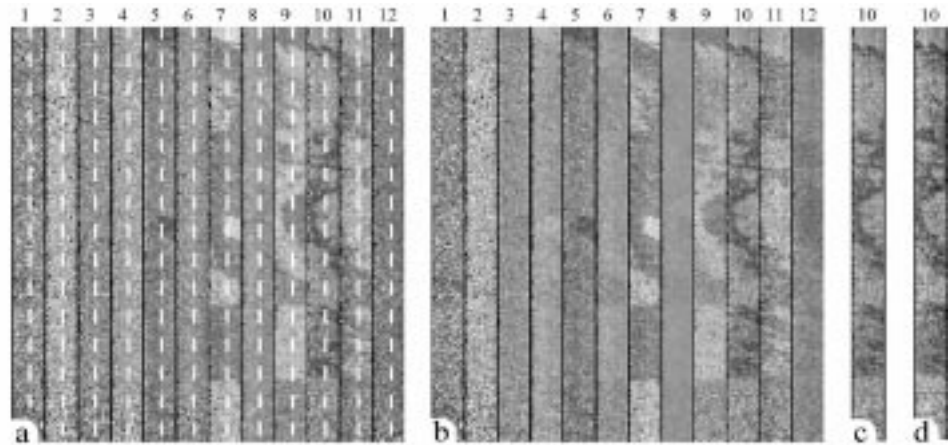


Figure 1. Results of the PCA multiwavelength image compression (a, b) and the whole-band interpolation (c, d). *Data courtesy of Yu. Shkuratov, Astronomical Observatory of Kharkiv University.*

#### 4. Implementation

For the initial training of the principal components a sample multiwavelength scan, indicated with the dashed lines in each of the 12 bands on Figure 1a, was arbitrarily chosen to produce the calibration. The eigenvectors were obtained using this scan containing 250 out of the total 4,750 spectra. Then the entire multispectral image set described above was compressed through encoding the data with the eigenvalues derived from the least-squares fits using the calibration eigenvectors. Each spectrum was processed independently from the others to simulate the real data recording process and without storing all the image in memory.

We found the six principal components providing the compression rate of 1:2, which are able to represent simply by the linear combinations all the features of the original data with the differences not exceeding the *rms* error in each channel. The restored multiwavelength image is shown on Figure 1b.

The calibration eigenvectors allow interpolation for the pixels or even whole spectral bands, which are affected by the impulse noise or other errors. To demonstrate this possibility we have deliberately excluded one channel in the original data set from consideration (shown on Figure 1c) and performed the data encoding with the six principal components. Image restoration made in the ordinary manner proves such a knowledge-based interpolation to be ideal for the remote sensing imagery (see the interpolated single-channel image on Figure 1d).

#### 5. Conclusions

PCA is shown to be applicable to on-board multispectral image compression allowing the incremental compression preceded by the preliminary calibration

of the principal components with fairly short learning times. This calibration is stable as it leans on a more varied data library and can be apparently performed on the basis of the laboratory spectra as well.

It is important to note that this type of compression is practically lossless at its reasonably high rates due to the PCA's ability to discard first and foremost the noise and redundant correlations from the data. It should also be noted that compression is better for the higher band numbers where the method retains its robustness and accuracy, our other experiments with hyperspectral land satellite images show that. The compression rate can be further increased by subsequent application of a spatial decorrelating techniques such as JPEG or DCT.

In addition to the reduction of the data dimensions PCA can be used for automatic correction of the impulse noise due to its unique interpolation abilities. The principal components so obtained are characterized by higher informational content than the initial spectra and can be directly used for various data interpretation tasks (see, for example, Shkuratov et al. 1996; Smith et al. 1985; Vasilyev 1996).

**Acknowledgments.** The author thanks the ADASS VII Organizing Committee for offering him full financial support to attend the Conference.

## References

- Birk, R. J. & McCord, T. B. 1994, *IEEE Aerospace and Electronics Systems Magazine*, 9, 26
- Genderon, R. G. & Goddard, M.C. 1966, *Photogr. Sci. Engng*, 10, 77
- Ksanfomaliti, L.V. 1995, *Solar System Research*, 28, 379
- Ksanfomaliti, L. V., Petrova, E. V., Chesalin, L. S., et al. 1995, *Solar System Research*, 28, 525
- Liu-Yue Wang & Oja, E. 1993, *Proc. 8th Scandinavian Conf. on Image Analysis*, 2, 1317
- Malinowski, E. R. 1977, *Annal. Chem.*, 49, 612
- Rao, A. K. & Bhargava, S. 1996, *IEEE Transactions on Geoscience and Remote Sensing*, 34, 385
- Roger, R. E. & Cavenor, M. C. 1996, *IEEE Transactions on Image Processing*, 5, 713
- Shkuratov, Yu. G., Kreslavskii, M. A., Ksanfomaliti, L. V., et al. 1996, *Astron. Vestnik*, 30, 165
- Smith, M. O., Johnson, P. E. & Adams, J. B. 1985, *Proc. Lunar Sci. Conf. XV. Houston. LPI*, 797
- Vasilyev, S. V. 1996, Ph.D. Thesis, Kharkiv St. Univ., Ukraine
- Vasilyev, S. 1997, in *ASP Conf. Ser.*, Vol. 125, *Astronomical Data Analysis Software and Systems VI*, ed. Gareth Hunt, & H.E. Payne (San Francisco: ASP), 155