SYNERGY PUBLISHING PTE. LTD.

Modern Electronic Technology

http://www.advmodoncolres.sg/index.php/amor/index

Application of PCA Numalgorithm in Remote Sensing Image Processing

Hong Dai*

School of Computer and Information Engineering, Shanghai Second Polytechnic University, Shanghai, 201209, China

ARTICLE INFO

Article history
Received: 5 February2023
Revised: 8 March 2023
Accepted: 10 March 2023
Published Online: 19 April 2023

Keywords:
PCA numerical algorithm
Remote sensing image processing
Multi-spectral image

ABSTRACT

A numerical algorithm of principal component analysis (PCA) is proposed and its application in remote sensing image processing is introduced: (1) Multispectral image compression; (2) Multi-spectral image noise cancellation; (3) Information fusion of multi-spectral images and spot panchromatic images. The software experiments verify and evaluate the effectiveness and accuracy of the proposed algorithm.

1. Introduction

Image processing is a technology of enhancing, pressing, correcting and identifying the images obtained by space remote sensing device and earth resource exploration platform to extract thematic information. It is a basic part of the whole remote sensing observation process [1]. Principle Component Analysis (PCA), also known as K-L transformation, is a multi-dimensional orthogonal transformation based on statistical characteristics. It is the best transform to remove the correlation between data, but its operation amount is large. This paper proposes a fast PCA numerical algorithm implemented in the computer, and realizes its application in the remote sensing image processing from three aspects through the software experiment: (1) Multi-spectral image compression. (2) Multi-spectral image noise cancellation. (3) Information fusion of multi-spectral images and spot panchromatic images.

2. The PCA Numerical Algorithm

Let an image sequence have n pictures, each graph size is k = M N (M-number of rows; N-number of columns), stack these images, and store the pixel value of each graph in a one-dimensional array, then the image sequence is represented by a two-dimensional array X, and an n-dimensional column vector X_i represents the set of all pixel values on position i, i.e.:

$$\boldsymbol{X_i} = [\mathbf{x}_{1i}, x_{2i}, \ldots, x_{ni}]^T \quad (1)$$

is

$$X = [\, X_1 \ldots \boldsymbol{X_k}]$$

The steps of transform X with PCA numerical algorithm are as follows:

(1) Calculates the mean vector m of X_X and the covari-

*Corresponding Author:
Hong Dai,
School of Computer and Information Engineering, Shanghai Second Polytechnic University, Shanghai, 201209, China;
Email: daihong419@163.com

ance matrix, C_X.

The mean vector of X is:

$$\boldsymbol{m}_X = E[X] = \frac{1}{k}\sum_{i=1}^{k} = X_i \tag{2}$$

The covariance matrix of X is:

$$\boldsymbol{C}_X = E[(X - m_X)(X - m_X)^T] = \frac{1}{k}\sum_{i=1}^{k}(X_i - \mathbf{m}_X)$$

$$(X_i - \mathbf{m}_X)^T = \begin{bmatrix} C_1 & C_2 & \dots & C_{1n} \\ C_2 & C_2 & \dots & C_{2n} \\ \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & \dots & C_n \end{bmatrix} \tag{3}$$

among C_{ij} (i,j=1,2,…n) reflects the correlation between pixels i and j. If the different pixels are unrelated, then C_X for diagonal arrays. C_X is the real symmetric array, hence the orthogonal matrix A, making C_X hierarchization, i.e.:

$$\Lambda = A^T \boldsymbol{C}_X A = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \lambda_n \end{bmatrix} \tag{4}$$

where the element on the diagonal is a C_X. The eigenvalue of, and sets the $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, A for C_X eigenvector matrices corresponding to these eigenvalues.

(2) Principal component transformation [2]

Let the matrix y be:

$$\boldsymbol{y} = \boldsymbol{A}\ (\boldsymbol{X} - \boldsymbol{m}_X) \tag{5}$$

Formula (5) is the principal component transformation of X. It tiably that the covariance matrix of y

$$\boldsymbol{C}_y = E[(y - m_y)(y - m_y)^T] = AC_X\boldsymbol{A}^T \tag{6}$$

Certificate: The mean value of y: $m_y = E[A\ (X - m_X)] = A\ E[X] - Am_X = \boldsymbol{Am_X} - \boldsymbol{Am_X} = \boldsymbol{0}$

The covariance array of y:

$\boldsymbol{C}_y = E[(y - m_y)(y - m_y)^T] = E[yy^T] - m_y\boldsymbol{m}_y^T = E[yy^T] = E[A\ (X - m_X)(X - m_X)^T\boldsymbol{A}^T] = A\ C_X\boldsymbol{A}^T$. By proof.

It is thus clear that C_y is a diagonal array, and the element on its main diagonal is called C_X. The eigenvalues of y column vectors are not correlated with each other. The principal component transformation can therefore remove the correlation between the data.

(3) Numerical algorithm for finding the eigenvector matrix A

In order to find A in a fast iterative way on the computer, the ancient and elegant comparable algorithm of solving the eigenvalue and the eigenvector of the real symmetric matrix is adopted first. The solution process is: At each iteration, in C_XSelect the element with the largest absolute value, and set it in column q of line p (let q> p), using the n-order rotation matrix R corresponding to the position$^{(k)}$To C_XOrthogonal similarity transformation, i.e.

$$\boldsymbol{C}_X^{(k)} = [R^{(k)}]^T \boldsymbol{C}_X^{(k-1)} \boldsymbol{R}^{(k)} (k = 1,2,\ldots\ldots) \tag{7}$$

among $C_X^{(0)} = C_X$, $R^{(k)}$The element $r_{pp}^{(k)} = \cos\theta$, $r_{qq}^{(k)} = \cos\theta$, $r_{pq}^{(k)} = -\sin\theta$, $r_{qp}^{(k)} = \sin\theta$, $r_{ij}^{(k)} = 0(i, j \neq p,q)$, $r_{ii}^{(k)} = 1(i \neq p,q)$.

1) If $r_{pp}^{(k-1)} - r_{qq}^{(k-1)} = 0$, (i) When $r_{pp}^{(k-1)} > 0, \theta = \pi/4$. (ii) When $r_{pp}^{(k-1)} < 0$, $\theta = -\pi/4$.

2) If $r_{pp}^{(k-1)} - r_{qq}^{(k-1)} = 0$, which is determined by the following equation:

$$tg2\theta = \frac{2r_{pq}^{(k-1)}}{r_{pp}^{(k-1)} - r_{qq}^{(k-1)}} \quad (|\theta| < \pi/4) \tag{8}$$

Each passing orthogonal similarity transformation, C_X. The sum of diagonal element squares increases by $2r_{pq}^2$. Again and again to C_X. Transform according to equation (7) until $|r_{pq}^{(k)}| < \varepsilon$ (p =q, a small positive number), you can be C_X. Gradually transform, for the diagonal array. The element on the diagonal in this diagonal array is a C_X. The eigenvalues of, the successive orthogonal similarity transformation matrix $R^{(k)}$. The product is C_X. The eigenvector matrix of A:

$$\boldsymbol{A} = \boldsymbol{R}^{(1)} \ldots R^{(k)} \tag{9}$$

Because of the ancient and elegant comparable method, it takes a time consuming to find the largest absolute value of the off-diagonal element every time, using the "pass method" [3]. Improvement: First, calculate C_X the sum of off-diagonal squares and root are:

$$v_0 \left[\sum_{i<j} (\boldsymbol{C}_{ij})^2\right]^{1/2} = 2 \tag{10}$$

Set Level 1 $v_1 = v_0 In / n$, as described in C_XScan by line in non-diagonal elements, if $|C_{ij}| > v_1$. To use the rotation matrix $R^{(k)}$To C_XOrthogonal similarity transform, otherwise "pass" and undergo multiple scans until all off-diagonal elements are less than v_1; Set pass 2 $v_2 = v_1$For / n, the above procedure is repeated until all off-diagonal elements are less than v_2; By analogy, through a series of passes $v_2, v_3, \ldots, v_r$. Until you meet the requirements of: $v_r \leq \rho v_1$, where is the accuracy requirement. The clearance method accelerates the convergence process of the Jacobian method.

(4) Error analysis

Inverse transformation formula through the principal component transformation: $X = A^T\boldsymbol{y} + \boldsymbol{m}_X$. Accurately rebuild the X. To compress the redundant information in this n image, then the m eigenvector A corresponding to the first m largest eigenvalues of y is taken$_1 \sim A_m$. To approximately reconstruct X, because $A_1 \sim A_m$. The main information in X is concentrated. The reconstruction formula is:

DOI: https://doi.org/10.26549/met.v7i1.12317

$$\boldsymbol{X_s}=\mathrm{y}\sum_{i=1}^{m}{}_i\boldsymbol{A}_i+\mathrm{m}_X \quad (11)$$

Then X and X_s The total mean squared error ems between is:

$$\mathrm{e\ ms}=E[(X-X_s)(X-X_s)^{T]}=\sum_{j=m+1}^{n}\lambda_j=\frac{\mathbf{1}}{\boldsymbol{MN}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{M}}\sum_{\boldsymbol{j=1}}^{\boldsymbol{N}}(\boldsymbol{X}-\boldsymbol{X_s})^{\mathbf{2}} \quad (12)$$

The number of feature vectors, m, is selected according to the error requirement.

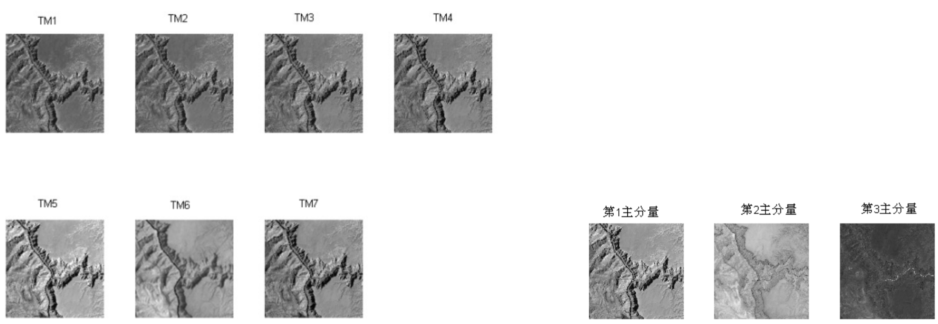
3. PCA Numerical Algorithm for Remote Sensing Image Compression

Remote sensing multi-spectral images have the characteristics of large data volume and high dimension, and massive data needs huge storage space [3]. The TM (thematic mapping instrument) multispectral images obtained by Landsat (Landsat) include seven bands: TM1 (visible blue light), TM2 (visible green light), TM3 (visible red light), TM4 (near infrared short wave), TM5 (near infrared medium wave), TM6 (thermal infrared) and TM7 (near infrared long wave) [2]. The correlation between the individual band images is very high, thus offering the possibility for image compression.

PCA numerical algorithm image compression experimental environment is a 256M memory of the Pentium computer and Matlab7.0 software. Data were obtained from a set of images from the Landsat satellite TM1 to TM7 band in the Colorado River, USA. Each drawing measures it at 512,512. Figure 1(a) was transformed by the PCA numerical algorithm. The first 3 principal component images are shown in Figure 1(b). The front m (m =1~7) main component images to approximately reconstruct the original image sequence. The proportion, compression ratio and reconstruction errors of their ground information are shown in Table 1, C_y the seven feature values are: $_1$=4889.90, λ_2=164.74, λ_3= 95.24, λ_4=49.08, λ_5=19.61, λ_6=10.57, λ_7=5.00. Visible that the first 3 principal component images already include:

$$(\lambda_1+\lambda_2+\lambda_3)/\sum_{\boldsymbol{i=1}}^{\mathbf{7}}\boldsymbol{\lambda_i}$$ = has 98.39% of the ground object information and the data volume is compressed to 42.9%. A 77 matrix C was solved in the experiments by Jacobi clearance$_X$The eigenvalues and eigenvectors, in the accuracy requirements of =10, respectively^{-2}, 10^{-3}And 10^{-4}. Next, the number of iterations was 4, 5, and 6 times, respectively.

And its principal component images: (a) TM1~7 band images (b) Top 3 primary component images).



(a) (b)

Figure 1. A TM multispectral image of the Colorado River

Table 1. Error analysis of image compression in the PCA algorithm.

Top m principal component images m	The portion of included ground information (%)	reduction ratio (Total bits of image / bits of compressed image)	ems (Total mean-squared error between the original image and the reconstructed image)
1	93.42	7.00	344.24
2	96.57	3.50	179.50
3	98.39	2.33	84.26
4	99.33	1.75	35.18
5	99.70	1.40	15.57
6	99.90	1.17	5.00
7	100	1	2.41×10^{-28}

4. PCA Numerical Algorithm for Remote Sensing Image Noise Cancellation

In remote sensing imaging in the process of imaging and image transmission have the influence of noise, CCD (charge coupling device) noise is the most important in the process of imaging noise, filter the DC component of CCD noise and eliminate zero response offset and response inconsistency, can be considered the remaining CCD noise for the superposition of Gaussian noise and salt and pepper noise [4]. The pepper noise can be eliminated by median filtering, and the remaining Gaussian random noise will reduce the performance of almost all compression algorithms. The probability density function of the Gaussian random variable z is given by the following equation:

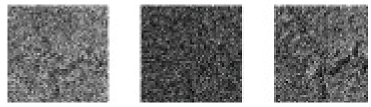
The p (z) $= \frac{1}{\sqrt{2\pi\sigma}} e^{-(z-\mu)^2/2\sigma^2}$ (13)

where z represents the gray value, the mean of z, is the standard deviation of z. Because Gaussian noise band and image band overlap, noise cannot be eliminated with traditional selective filter. As can be seen from Table 1, the amount of information contained in the remote sensing images after the main component transformation is gradually reduced, while the last three main component images contain very little ground object information. Since the Gaussian random noise components are uncorrelated (independent), the noise is highlighted when the information is reduced, so the transformation can separate the noise in the image.

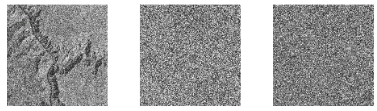
The experimental environment in this section is the same as in Section 2, and the data are still derived from the same set of images in the Colorado River T M 1 to T M 7 band, but all include a mean =0, and the normalized variance^{2}Gaussian noise =0.09, the first 3 images see Figure 3(a), the seven images, the first 3 main component images see Figure 2(b), obviously, only the first main component image contains more ground information in the next two subpictures almost only noise, so only take the first main component image to reconstruct the original image sequence, obviously reconstructed image (Figure 2(c) listed 3) Gaussian noise is reduced. Using the mean-square signal-to-noise ratio $(SNR)_{ms}$). To measure the effect of noise elimination, the formula is:

$$SNR_{ms} = \frac{\sum_{i=1}^{7} \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} \hat{f}(i,x,y)^2}{\sum_{i=1}^{7} \sum_{x=0}^{N-1}\sum_{y=0}^{M-1} [f(i,x,y) - \hat{f}(i,x,y)]^2} \quad (14)$$

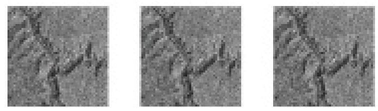
$\hat{f}$ where f is the original image; it is the reconstructed image. The SNR of the original noisy image was the SNR_0=4.31, and the SNR of the reconstructed images is the SNR_1=30.07, together with the SNR_0. Compared with the 6.97 times higher, the PCA numerical algorithm has a strong ability to eliminate Gaussian noise.



(a)



(b)



(c)

Figure 2. Remote sensing image noise cancellation by PCA numerical algorithm: (a) Multispectral images of the Colorado River with Gaussian noise; (b) Primary component image; (c) Denoised image only 3 pictures are listed.

5. The Numerical Algorithm of PCA Is a Remote Sensing Image Fusion

Remote sensing applications sometimes require both high spatial resolution and spectral resolution. However, imaging sensors of a single resource satellite are difficult to provide such data.

For example, the Landsat satellite can provide high spectral TM images with low resolution but low spatial resolution, while the Spot (Earth Observation Experimental System) satellite provides high spatial resolution panchromatic images with low spectral resolution. The key to solve this problem is the image fusion technology, which is to synthesize the scanning images of multiple sensors in the same area, and become a new image through the complementary information set, so as to achieve the effect of improving the spatial resolution while retaining the multispectral information [5].

PCA numerical algorithm image fusion experimental environment is the same as in section 2, the data comes from the spot TM image (Figure 3a) and the same area image (Figure 3b), Figure 3a, using PCA numerical algorithm, using the spot contrast linear stretch of the first main color image in Figure 3c, so it also maintains the color features of the color image high resolution, the color of the fusion image and the original image.

DOI: https://doi.org/10.26549/met.v7i1.12317

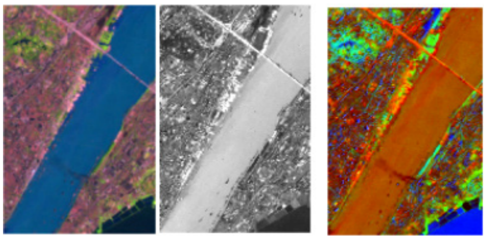


Figure 3. Remote sensing image fusion by PCA numerical algorithm: (a) TM multi-spectral visible light image; (b) Spot panchromatic image; (c) Fused image.

6. Conclusions

This paper presents the PCA numerical algorithm, available quickly in a computer, for compression, noise cancellation and fusion of remote sensing images, and verifies the effectiveness and accuracy of the algorithm with experiments.

References

[1] Wei, Y.Ch., Tang, G.A., Yang, X., et al., 2001. Remote sensing digital Image processing tutorial. Beijing: Science Press. pp. 1-30.

[2] Gonzalez, R.C., Woods, R.E., 2003. Digital image processing (Version 2). Beijing: Electronic Industry Press. pp. 550-560.

[3] Zhang, J., Lu, Y.H., Liu, Q.Y., 2006. Application of an improved embedded wavelet algorithm in remote sensing image compression. Microcomputer Information. (3-2), 79-81.

[4] Zhong, W.B., Ning, Sh.N., Jin, S.Z., et al., 2004. Remote sensing imaging noise analysis and the PCNN-based filtering method. Journal of Coal. (4), 418-421.

[5] Gao, Sh.Ch., 2006. VISUAL C + + practice and improvement: Digital image processing and engineering application. Beijing: China Railway Press. pp. 326-342.

 DOI: https://doi.org/10.26549/met.v7i1.12317