Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Image Compression and Reconstruction Based on PCA

To cite this article: Linglong Tan *et al* 2021 *J. Phys.: Conf. Ser.* **1944** 012021

View the article online for updates and enhancements.

You may also like

- A New Proposed Method For A Statistical Rules-Based Digital Image Compression
Khadhim Mahdi Hashim, Salwa Shakir Baawi and Bushra Kamil Hilal
- MRI image compression using multiple wavelets at different levels of discrete wavelets transform
Prakash S Narayana and A M Khan
- Developed JPEG Algorithm Applied in Image Compression
Abdullah A. Hussain, Ghadah K. AL-Khafaji and Mohammed M. Siddeq


245th ECS Meeting
San Francisco, CA
May 26–30, 2024

PRiME 2024
Honolulu, Hawaii
October 6–11, 2024

Bringing together industry, researchers, and government across 50 symposia in electrochemistry and solid state science and technology

Learn more about ECS Meetings at
http://www.electrochem.org/upcoming-meetings

Save the Dates for future ECS Meetings!

This content was downloaded from IP address 202.12.83.168 on 06/06/2023 at 16:57

Image Compression and Reconstruction Based on PCA

Linglong Tan[1*], Fengzhi Wu[2]and Weilong Li[3]

[1]Electronic engineering college, Anhui Xinhua University, Hefei, 230088, China
[2]Ashikaga University,
information Science and Manufacturing Engineering,Tochigi,Ashikaga-shi,326-8558
[3]Anhui Academy of Coal Science, Hefei,230001,China

[*]E-mail: tanlinglong@axhu.edu.cn;

Abstract. In view of the disadvantages of large image space, high dimension of feature representation and large storage, this paper uses principal component analysis to compress the data, which can effectively reduce the loss of information, reduce the dimension of data and extract features in all aspects of image compression. In this paper, based on the analysis of PCA algorithm dimension reduction, reconstruction, feature extraction principle, the PCA algorithm is applied to image compression and reconstruction. Through the principal component analysis of sample variables, the feature vector of sample variance is calculated, and its principal component is extracted, and the contribution rate is calculated to realize image PCA compression. By extracting different eigenvalues from the original image and calculating its compression ratio and contribution rate, we can see that the clarity of the image is positively correlated with the value of the eigenvalue, and the larger the value of the eigenvalue, the higher the contribution rate. PCA based image processing technology is simple to use, high compression rate and high quality of image reconstruction, which is unmatched by many other methods.

1. Introduction

Principal components analysis (PCA) is a method of mathematical transformation [1, 2]. Its main purpose is to use the idea of dimensionality reduction to integrate multiple information components into several components to achieve ‘compression’ of information. According to statistical principles, Principal Component Analysis can be regarded as a simplified data set technique, which is transformed according to linear rules, and this transformation can map the data to a new coordinate system. The first largest variance of any data projection is on the first coordinate (called the first principal component), the second largest variance is on the second coordinate, and so on [3]. The amount of storage and transmission of images is huge, and at the same time it creates pressure on the memory and rate of information storage and transmission. The use of PCA technology can effectively reduce the image dimension. Through the feature extraction of PCA technology, the most effective information part of the image can be obtained. Filter the invalid part of the image, thereby compress the image, and reconstruct the image that meets the actual needs. Principal component analysis is a statistical method that uses the idea of dimensionality reduction and uses several uncorrelated principal components to reflect most of the information of the original variable. Principal component analysis plays an important role in various scientific and engineering fields, because it can reduce the dimensionality of data to achieve the purpose of processing information and minimize the amount of error between the extracted components and the original data.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Basic principles of PCA

The basic principle of principal component analysis is to select a coordinate axis projection with a large variance for the sample points, and use this method to reduce the dimensionality of the data without any obvious loss of information. The sample matrix $X_{m\times n}(m > n)$is composed by centralizing the given data $\{xi\}_{i=1}^{m}$, at this time $x_i \in R^n$,and

$$\sum_{i=1}^{m} xi=0 \tag{1}$$

If the given image data A has not been centrally processed, then the formula (1) does not hold, and the number of processing A should be standardized next. The formula can be expressed as:

$$X_{ij}=\frac{A_{ij}}{S_j}-\frac{\bar{A}_j}{S_i} \tag{2}$$

Thus the mean of the sample and the standard deviation of the sample are:

$$\bar{A}_j=\frac{1}{m}\sum_{i=1}^{m} A_{ij} \qquad S_j=\sqrt{\frac{1}{m-1}\sum_{i=1}^{m}\left(A_{ij}-\bar{A}_j\right)^2} \tag{3}$$

available $X=\left(x_{ij}\right)_{m\times n}$

Through formula (3), the previously input data vector X_i is transformed into a new data vector by PCA

$$S_i = U^T x_i \tag{4}$$

In formula (4): U is the orthogonal matrix of $n \times n$, and the i-th eigenvector of the sample covariance matrix is its i-th column U_i.

$$C = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^T \tag{5}$$

In other words, PCA answers the following intrinsic questions:

$$\lambda_i U_i = C u_i \quad i = 1,\cdots,n \tag{6}$$

Here $\lambda$is actually one of the eigenvalues of C, and its corresponding eigenvector is U_i. If only the previous p eigenvectors are used as eigenvectors, the conversion of PCA can be expressed by the following formula:

$$S = U^T X \tag{7}$$

In formula 7, S is the main component, and the maximum eigenvector u corresponding to the maximum eigenvalue λ is called the first principal component[4,5]. The second principal component is also the eigenvector corresponding to the second largest eigenvalue. The variance data point will produce the second largest change in this specific direction. Here, this eigenvector and the first eigenvector are in principle the same orthogonal relationship. The PCA transformation matrix U transforms the kernel matrix according to the size of the eigenvalue λ . Since its main effective part is concentrated in the coefficient with a large eigenvalue λ , the first k principal components with the larger eigenvalue are used to approximate S , here $k < (n \times n)$, that is, if the coefficient with the smaller feature value$\lambda$is eliminated, the image will maintain close to the original quality within a certain range. The new transformation matrix $U_k$is a combination of k eigenvectors corresponding to the previous k largest eigenvalues, and a new transformation is performed on U_k .

$$S_k = U_k X, U_k = \begin{bmatrix} u_1^T \\ u^T{}_2 \\ \cdots\cdots \\ u_k^T \end{bmatrix}_{k\times n\times n} \tag{8}$$

The formula (8) represents the principal component of the image. The above theory is the basic principle of PCA.

PCA can play a major role in data dimensionality reduction. For vectors with higher dimensions, a projection matrix can be obtained by using the PCA method. This matrix has k-dimensional features (low relative dimensionality), and this projection matrix can complete the dimensionality reduction function from high to low dimensions.Each dimension of the new low-dimensional feature is represented as an orthogonal relationship, and the feature vector also maintains an orthogonal form. After dimensionality reduction, the redundant components of the feature vector are effectively reduced, and its correlation is greatly reduced, which reflects the basic characteristics of its features to a certain extent, and will not have a large number of repetitive effects in subsequent classification predictions.

Input an image of size $w \times h$, and its columns are connected to form a $M = w \times h$ dimensional column vector. If the number of images is N, the image matrix of the column vector X of the j -th image of X_j dimension is composed of N images arranged together and recombined. Then the overall covariance matrix:

$$\sum = \frac{1}{N}\sum_{i=1}^{N}(X_i -)(X_i - u)^T \tag{9}$$

Among them, u is the average image vector of the sample set images,$u = \frac{1}{N}\sum_{i=1}^{N} X_i$

Then algorithms such as QR or SVD can be used to solve the eigenvalues of the matrix and the corresponding eigenvectors. The first L eigenvectors are extracted to form its projection matrix $Eig = (u_1, u_2, \cdots, u_L)$, and the value of L can be determined by the cumulative contribution rate of the eigenvalues:

$$\alpha \leq \frac{\sum_{i=1}^{L}\lambda_i}{\sum_{i=1}^{wh}\lambda_i} \tag{10}$$

Eig is the projection matrix. The original high-dimensional $(wh \times 1)$ image column X_i is reduced by the projection matrix as follows:

$$\mathrm{Feature}_i = \mathrm{Eig}^T \cdot X_i \tag{11}$$

$\mathrm{Feature}_i$ is the column vector of $L \times 1$. In this way, the dimensionality reduction processing of the original vector is completed, the useful feature part of the original vector is extracted, and the internal effective content of the original vector is better preserved.

The main content of this method is to extract the main features of the information. The key step is to select a suitable classification algorithm to classify the extracted image features [6]. The use of PCA for feature extraction is the most common method. Many applications in space dimensionality reduction are based on the feature extraction of PCA. Through feature extraction, the data is classified into different categories. This is exactly what the field of image processing is deeply explored. section. For example, this method is widely used in fields such as face recognition, image processing, and engineering data extraction.

3. PCA image reconstruction principle

The principal component $S_k$in PCA reduces the $n \times n - k$ dimension relative to the dimension of S, and the reverse transformation is performed to obtain the dimensionality reduction reconstruction value $\hat{A}$ of the original image A:

$$\hat{A} = U_k^T S_k \tag{12}$$

In the above formula, the image reconstructed by PCA is the estimated value $\hat{A}$ of the original image A, and there is also an inevitable error here. The following mathematical formula can express the error value between A and $\hat{A}$:

$$e_{ms} = \sum_{i=1}^{n\times n}\lambda_i - \sum_{i=1}^{k}\lambda_i = \sum_{i=1}^{n\times n}\lambda_i \tag{13}$$

The formula (12) shows that if $k = n \times n$ transforms all eigenvectors, there will be no error.If k eigenvectors (with the largest eigenvalues) are selected to form the transformation matrix U_k, and the image is reconstructed by dimensionality reduction, its mean square error can be reduced to a negligible degree, so the principal component analysis method achieves better results in image compression and reconstruction.Then the obtained data with n indicators can be expressed by an $m \times n$-dimensional data matrix, and the expression formula is as follows:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \quad (14)$$

Next, the matrix A is standardized, and the matrix index component in formula (13) is standardized, and the processed matrix can be obtained by formula $X_{ij}=\frac{A_{ij}}{S_j} - \frac{\bar{A}_j}{S_i}$ as $X = (x_{ij})_{m\times n}$.

Calculate the correlation coefficient matrix of the sample matrix by formula (12):

$$R = \frac{1}{m-1} X^T \cdot X = (r_{ij})_{m\times n} \quad (15)$$

Using the Schmidt orthogonalization method, the unit orthogonalization eigenvectors are used to obtain $a_1, \cdots, a_1$, and the cumulative contribution rate $B_1, \ldots, B_n$ of its eigenvalues is calculated. According to the defined extraction efficiency p, if $B_t \geq p$, then extract t principal components $a_1, \ldots, a_t$. Then calculate the projection $Y = X \cdot a$ of the normalized sample data X on the extracted feature vector, where $a = (a_1 \cdots a_t)$.The obtained Y is the data after feature extraction, which is the data after data dimensionality reduction [7].

From equation (14), the reconstructed image $\hat{X}$ can be obtained, and the formula for inverse transformation is:

$$\hat{X}=X^T Y \quad (16)$$

4. Simulation Realization of Image Compression and Reconstruction Based on PCA

4.1 Algorithm analysis

The principal component analysis based on sample variables is realized through the function pcasample. First, the eigenvector of the sample variance is calculated, the first p principal components are extracted, the score matrix is calculated, and the contribution rate is calculated finally.The PCA compression of the image is realized by the function pcaimage. The color image is first converted into a gray image, and then the image array is converted into a sample matrix, the number of samples and variables are selected, and finally the image is reconstructed with the extracted first P main components.

4.2 Simulation results and analysis

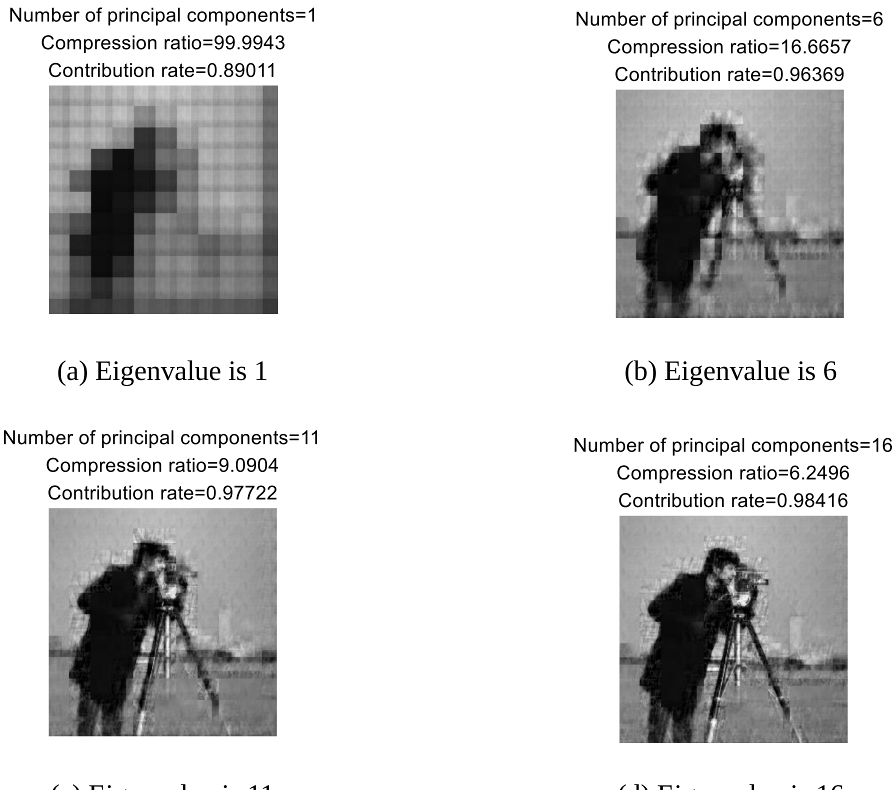
Figure 1(a) is the original image of the simulation. The original image is compressed and reconstructed using PCA image processing technology, and the eigenvalues are 1, 4, 8, and 16 for simulation experiments.The simulation results are shown in Figure 2(a), Figure 2(b), Figure 2(c), and Figure 2(d).Figure 2(a) is the result of compressing and reconstructing the information source image. Its feature value is 1, the compression ratio is 99.9943, and the contribution rate is 0.89011.Figure 2(b) takes the characteristic value of 6, the compression ratio of 16.6657, and the contribution rate of 0.96369.Figure 2(c) has a characteristic value of 11, a compression ratio of 9.0904, and a contribution rate of 0.97722.Figure 2(d) has a characteristic value of 16, a compression ratio of 6.2496, and a contribution rate of 0.98416.It can be seen from Figure 2 that the sharpness of the image is positively correlated with the value of the feature value, and the larger the value of the feature value, the higher its contribution rate. In daily needs, image compression and reconstruction techniques based on principal component analysis can be adjusted according to actual conditions.

In the simulation experiment of image compression and reconstruction based on principal component analysis, the experimental principles can be understood more clearly. With the continuous development of electronic media information technology, the content carried in image multimedia has also become larger. Usually the image is composed of a group of highly correlated with each other, the

Journal of Physics: Conference Series **1944** (2021) 012021 doi:10.1088/1742-6596/1944/1/012021

amount of data is large, and it takes up a lot of storage space [8]. The use of principal component analysis to process images has natural advantages for storing and transmitting these data. Therefore, the degree of interest in image compression technology is increasing. PCA-based image processing technology is simple to use, with high compression rate and high image reconstruction quality, which is unmatched by many other methods.

Figure 1. Original image



(a) Eigenvalue is 1 (b) Eigenvalue is 6

(c) Eigenvalue is 11 (d) Eigenvalue is 16

Figure 2. Image compression simulation example diagram

5. Summary

PCA is more effective for reducing the dimensionality of the data. There will be no large mean square error between the extracted components and the original data. It is commonly used in feature extraction and data compression. With the further development of image processing and information technology based on electronic media, a large amount of digital information is contained in widely distributed image media. In order to make it more convenient to store and transmit image data in the media, people are paying more attention to image compression technology. It can be seen from the experiment that this processing method is more reliable and convenient, can complete image compression and reconstruction efficiently, and can recover different data images according to the number of its principal components, which can meet the needs of different levels of image reconstruction .

Acknowledgments
This work was supported by the Quality Engineering Project No.2015zy073, Scientific Research Project of Anhui Natural Science Research Project No. KJ2020A0785 and the Research Team Project kytd201904.

References
[1] Microscopy; Findings in Microscopy Reported from Changchun University of Science and Technology (A Medical Image Compression and Reconstruction Method Based On Improved Principal Components Analysis) [J]. Science Letter, 2020.
[2] Microscopy; Findings in Microscopy Reported from Changchun University of Science and Technology (A Medical Image Compression and Reconstruction Method Based On Improved Principal Components Analysis) [J]. Science Letter, 2020.
[3] Zhiqing Lu,Jin Tang. Multiple Reconstruction Compression Framework based on PNG Image[J]. International Journal of Computational Science and Information Technology,2019,7(4).
[4] Sheng CANG,A-chuan WANG. Research on Compression Perceptual Hyperspectral Image Reconstruction Based on GISMT[A]. Advanced Science and Industry Research Center.Proceedings of 2018 International Conference on Computer, Communications and Mechatronics Engineering(CCME 2018)[C].Advanced Science and Industry Research Center:Science and Engineering Research Center,2018:6.
[5] Yanwen Chong,Weiling Zheng,Haonan Li,Zhixi Qiao,Shaoming Pan. Hyperspectral Image Compression and Reconstruction Based on Block-Sparse Dictionary Learning[J]. Journal of the Indian Society of Remote Sensing,2018,46(7).
[6] Information Technology - Information Systems; New Data from Changchun University of Science and Technology Illuminate Findings in Information Systems (Improved Sliding Shapes for Instance Segmentation of Amodal 3d Object)[J] Computers Networks & Communications, 2019
[7] T.B. Knudsen Integrative Modeling and AOP-based Simulation of Developmental Toxicity[J] Toxicology Letters, 2018, 295
[8] Rolf Hilgenfeld Enterovirus 3C Proteases: Structure-based Discovery of Inhibitors[J] Antiviral Research, 2010, 86(1)