Wasit Journal of Engineering Sciences

Journal homepage: https://ejuow.uowasit.edu.iq

Digital-Image Dimension Reduction Via Analysis of Principal component

Rusul F. Zakur[1a] Ismail Sh. Hburi[1] Hasaneen Flaih[1] Mays M.Taher[1] and Hasan F. Khazaal[1]

Affiliations
[1]Electrical Engineering Department, Wasit University, Wasit, Iraq

Correspondence
isharhan@uowasit.edu.iq

Received
05-March-2022
Revised1
17- June -2022
Accepted
18-July-2022
Doi: 10.31185/ejuow.Vol10.Iss2.304

Abstract

A high-resolution image takes up more storage space, in addition to the data transit difficulty. The Analysis of Principal Component, or PCA for a brief notation, is a mathematical approach utilized to lessen the data dimensionality. It extracts the main pattern of a linear system using the factoring matrices technique. The objectives of this paper are to see how effective PCA is in reducing digital picture features and to investigate the (feature-reduced) images' quality on comparison with different values of the variance. As per the synthesizing of the initial research, the dimension or size reduction technique through the Analysis of Principal Component typically involves of 4-important steps: (1) picture-data normalizing (2) matrix of the covariance calculating using picture-data. (3) discovering the picture-data projection (with fewer number of features) to a new basis use the Single Value Decomposition technique (SVD) (4) determining the picture-data projection (with fewer number of characteristics) to a new basis. According to testing (500 picture) in language python and program anaconda3 results, the PCA approach considerably decreases the size of picture data while sustaining the original picture's fundamental properties. This approach reduced file size by 35.3 percent for the best feature lowered quality. The upload time of picture files through the Internet has substantially improved, particularly for mobile device downloads.

Keywords: image processing,PCA,SVD.

الخلاصة : تشغل الصورة عالية الدقة مساحة تخزين أكبر ، بالإضافة إلى صعوبة نقل البيانات) PCA .تحليل المكون الرئيسي) هو نهج رياضي يستخدم لتقليل أبعاد البيانات. يستخرج النمط الرئيسي لنظام خطي باستخدام تقنية مصفوفات التحليل. تتمثل أهداف هذه الورقة في معرفة مدى فعالية PCA في تقليل ميزات الصورة الرقمية ومقارنة جودة الصور ذات الميزات المختصرة بقيم تباين مختلفة. كنتيجة لتوليف البحث الأولي ، تتكون تقنية تقليل الأبعاد من خلال PCA عادةً من أربع خطوات مهمة: (١) تطبيع بيانات الصورة (٢) حساب مصفوفة التباين من بيانات الصورة (٣) استخدام تحليل القيمة الفردية (SVD) لاكتشاف إسقاط بيانات الصورة على أساس جديد مع عدد أقل من الميزات (٤) ابحث عن إسقاط بيانات الصورة على أساس جديد بخصائص أقل. وفقًا لنتائج الاختبارعلى ٥٠٠ صورة باستخدام لغة البايثون وبرنامج (anconda 3)، يقلل نهج PCA بشكل كبير من أبعاد بيانات الصورة مع الحفاظ على الخصائص الأساسية للصورة الأصلية. أدى هذا الأسلوب إلى تقليل حجم الملف بنسبة ٣٥.٣ بالمائة للحصول على أفضل ميزة ذات جودة منخفضة. تحسن وقت إرسال ملفات الصور عبر الإنترنت إلى حد كبير ، لا سيما بالنسبة لتنزيلات الأجهزة المحمولة

1.INTRODUCTION

In information technology, specifically in this advanced information era, the successful digital pictures transmission has become a key task or challenge [1]. Large numbers of image data are transferred through the Internet on daily basis. Meanwhile, the users' number of mobile phones devices that are connected to the Internet is growing rapidly. Pictures with high-resolution require more bits of information for transfer and storage, it would need a long time to display clearly on the screen of the browser. Given the significant bandwidth utilization of video streaming, having a suitable compression coding approach with a low bit - rate is important for efficient and rapid picture data transmission over the internet. Picture compression is based on decreasing the number of bits or information required to characterize a photo. Unnecessary extra data is removed from the picture's substance by compressing it (details). The 2-type of techniques for compression is lossless and lossy compression. Later approach reduces the size of the image file by removing extraneous data from the image content. This method will cause data to be lost during the process of compression. As a result, the final image will have a level of visual degradation that is fair.

Lossy type compression on the other hand has been extensively utilized on multimedia files, where the producing digital material quality remains untouched but the file size is significantly reduced. When a file is decompressed, lossless compression keeps the original file's quality. It is documented that lossless compression is used [2-6]. In this work, to remove the unnecessary portions that need a large amount of memory of digital photographs, the statistical technique is utilized for this task. In general, photos with high resolution are regarded to be large dimensional information. In a problem space, picture information is considered to be a matrix pixel of 2D. The pixel's color can be determined as an RGB bit value for a given pixel element. Accordingly, the Analysis of the Principal Component (PCA) is an efficient approach for minimizing the dimensional space that mathematical functions provide [7,8]. In addition, for a linear system, the SVD is an efficient approach that can be utilized to extract the basic system's properties [9]. The Analysis of the Principal Component has been commonly utilized to reduce the level of noisy signals in (DSP), picture recognition, and classification difficulties [14]. In this study, for different values of variance, the quality of photos with feature reduction is examined, and the PCA digital image compression technique is divided into four different categories. A pair of Internet connected devices (desktop and mobile set) will be used for transfer times' comparison of the original and resulting picture. Visual data in our work is supposed to be expressed by a linear function. Prior to compression, image data is normalized to confirm that all picture information falls within the specified range. The paper sections is:1.introduction which explain the main points,2.related work contain some studies and 3. Analysis of the Principal Component,4. Digital Image Compression Progression, section 5. Results and Discussion and the last section is 6.the conclusion.

2. RELATED WORK

Reference [15] utilized Analysis of the Principal Component to reduce digital pictures' size for healthcare application. More Analysis of the Principal Component aided digital picture reduction information may be found in [16-20]. Fast Fourier Transform, artificial neural approach, Discrete Cosine Transform, and Discrete Wavelet Transform, are state-of-the-art compression coding methods utilized in the JPEG standard [21-25]. FFT and DCT both compress images by transforming values of pixel (2-D matrix values) into a two-dimensional frequency domain waveform matrix. Then, in the frequency domain, the unnecessary components of frequency are eliminated., Sinusoids waveforms are utilized as the basis equation in Fourier series analysis. On the other hand, Wavelet analysis is relay on the use of orthonormal basis functions to break down data [26]. Before evaluating visual input, leverage a wavelet, DWT decomposes this input into several components' layers of frequency. The majority of the picture information is carried by components with low-frequency. As a result, components with low-frequency are frequently linked to visual patterns. The high-frequency component adds features to images. Picture information with a value less than a selected level or threshold is then rejected [27-29], and de-noising in high-frequency components is performed. After that, the filtered pictures are synthesized into lower-dimensional picture. The DWT compression technique is difficult when compared to other methods, when a correct threshold level is picked in the phase of de-noising. The DWT also has the ability to examine data in a number of scales and translations. Because of its flexibility, the wavelet transform can have an infinite number of different basis functions.

3.ANALYSIS OF THE PRINCIPAL COMPONENT

Analysis of the Principal Component is a statistical technique that uses total variance to determine the fundamental properties of a dispersed dataset 30. First, for a collection of distributed information (multi-variate) in a 2-D coordination system, Analysis of the Principal Component calculates the biggest original dataset's variations as shown in Figure (1). After that, this information is projected onto a different coordinate system axis, i.e., the U-V axis, and here the principal components are represented by U and V axes' directions i.e., the major orientations in which the information is going.

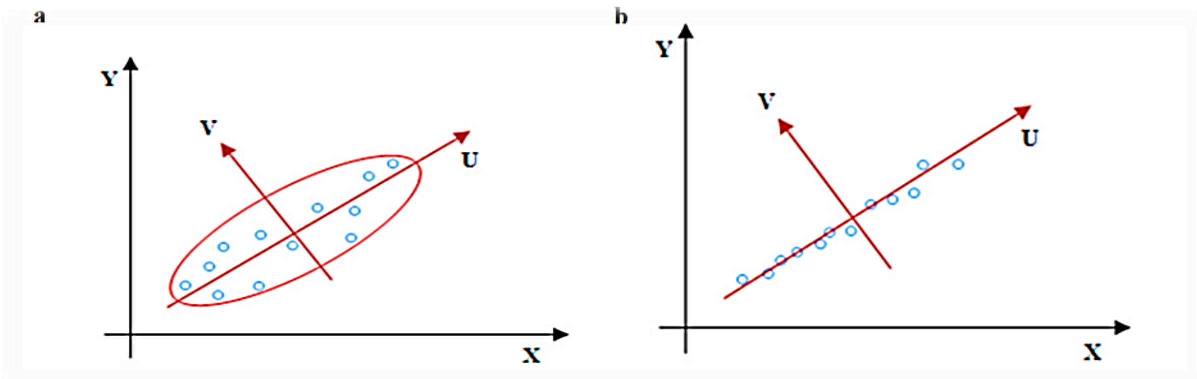

Figure1. Dimension compression via Analysis of the Principal Component, (a) altered system of axis; (b) eliminated variable v.

When all of the information on the V axis is extremely near to null, fig, 1-b demonstrates how we can employ only one variable i.e., var. U to express the information while neglecting the other variable i.e. var. V.

4. DIGITAL IMAGE COMPRESSION PROGRESSION

A 2-D matrix form set of pixels or dots can be utilized to express a digital picture as revealed in equation (1) [15]. When it comes to colored pictures, these dots are in floating-point format, whereas they are in discrete values in gray scaled images.

$$g(x,y) = [g(0,0)\ \cdots\ g(0,m-1)\ \vdots \ddots \vdots\ g(n-1,0)\ \cdots\ g(n-1,m-1)\ ] \tag{1}$$

where y and x are the picture pixel's coordination and g (x,y) is the image's matching gray level or color, according to the category of the value. The four main processes in the (PCA) [15] image dimension reduction technique are; normalizing the picture, determining the picture information covariance matrix, computing the covariance matrix eigenvalues and eigenvectors, and translating picture information into a different basis.

4.1 Normalizing the image

As a part of the PCA pre-processing stage, the image dataset is normalized. By deleting the picture dataset with their mean value as shown in equation (2), the dataset's discriminating power and the ratio of signal-to-noise may be enhanced [15].

$$G_{normalized}(x,y)=[g(0,0)\ \cdots\ g(0,m-1)\ \vdots \ddots \vdots\ g(n-1,0)\ \cdots\ g(n-1,m-1)\ ] \tag{2}$$

here [g(0,0) … g(0,m-1)] is the average value for y_1 through y_m (column vector).

4.2 Image data covariance matrix

The largest value of variance in the information set can be determined based on the covariance. The image's covariance-matrix, Cov(x,y), may be calculated using the following equation [15]:

$$Cov(x,y)=\frac{F_{normalized}\ (x,y)*F_{normalized}\ (x,y)T}{m-1} \tag{3}$$

Here m is the elements number.

4.3 Covariance-matrix eigenvalues and eigenvectors

The direction of the image information set (with the greatest variance) can be represented by the eigen vector (associated with highest eigen value). The important properties of the picture information are included within the eigenvector matrix [15]. The SVD expression can be employed to determine the eigenvalues and eigenvectors of the AA^T covariance matrix and may be expressed as below :

$$AA^T = Cov(x,y) = UD^2 U^T \quad (4)$$

The eigenvectors of (AA^T) are (U) , and the eigenvalues of (AA^T) are the singular values' square in Matrix D.

4.4 Different basis transformation of image information

At this stage, a new axis approach can be used to minimize the dimension of the original image collection by converting it into. Eq. 5 below and this new axis system is employed to project the original image using the eigenvectors produced by equation (4):

$$F_{transformed}(x,y) = U^T F_{normalized}(x,y) \quad (5)$$

Formalized (x,y) is the corrected original photo dataset, U is the eigen-vectors matrix and $(.)^T$ is the matrix transpose operation[15].

5. RESULTS AND DISCUSSION

The photo in Fig. 2 is a colorful display that has been converted to a gray color, and it is introduced in Fig. 3 after the process has been completed. The original image is compared to the feature-reduced image after PCA is applied. The feature-reduced images with variances of (0.97) and (0.95) are shown in Figures 3(b) and 3(c). It can be seen that the PCA approach assures the following attribute; the picture data of less important features have been removed, but the image's primary qualities have not changed. Instead, when compared to the original picture, the file sizes of Figure 3(b) and 3(c) have been decreased by 39% and 40%, respectively. The original photo and its feature-reduced image are shown in Figure (3 (variance = 0.97). The original image's dimension was lowered after the PCA procedure without sacrificing too much of the image's content. The image's file size has been decreased by 35.3 percent.

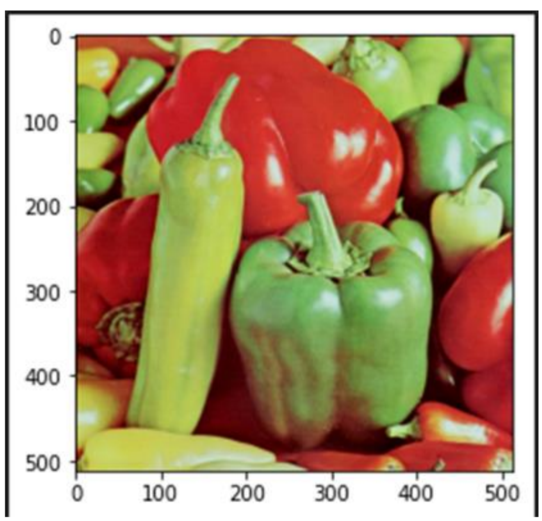

Figure 2: the main image

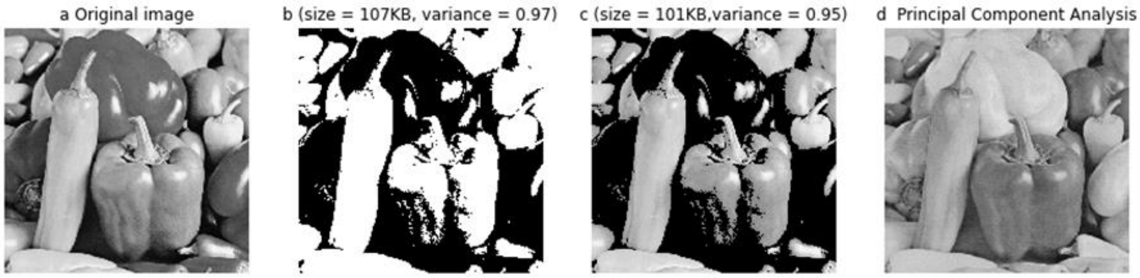

Figure 3: (a) The raw image i.e., original one with size of 175 Kbits. **(b)** processed image with variance of 0.97 and size of 107 Kbits (feature-decreased). **(c)** processed image with variance of 0.95 and size of 101 Kbits (feature-decreased)

6. CONCLUSIONS

This work investigates how PCA can be supportive in image's feature dropping for DSP systems. The principal components of the dataset may be discovered by computing the covariance matrix's eigenvectors and eigenvalues. These main components keep the original picture's feature factor information. A lossy compression approach is the use of PCA for image data reduction. . The variance value you select is critical since it influences image quality and file size. Although the size reduction of photo-file may save time during transmission (in uploading and downloading), the visual quality will suffer or degrade. Our research outcomes depict that this strategy saves a substantial amount of memory space and the time which is required for the transmission of photo files while preserving image quality. This technique, however, is restricted to uncorrelated or decoupled variables linear problem space. In future work, an alternate image size compression technique for dealing with correlated or coupled variables non-linear systems will be examined.

REFERENCES

1. Gelogo, Y. E., & Kim, T. H. (2013). Compressed images transmission issues and solutions. *Int J Comput Graph*, ***5***(1), 1-8.
2. Kokovin, V. A., Uvaysova, S. S., & Uvaysov, S. U. (2016, May). Lossless compression algorithm for use in telecommunication systems. In *2016 International Siberian Conference on Control and Communications (SIBCON)* (pp. 1-4). IEEE.
3. Lin, P. C., Pai, Y. H., Chiu, Y. H., Fang, S. Y., & Chen, C. C. P. (2016, March). Lossless compression algorithm based on dictionary coding for multiple e-beam direct write system. In *2016 Design, Automation & Test in Europe Conference & Exhibition* (pp. 285-288). IEEE.
4. Dzhagaryan, A., & Milenkovic, A. (2015, October). On effectiveness of lossless compression in transferring mHealth data files. In *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)* (pp. 665-668). IEEE.
5. Wandelt, S., Sun, X., & Zhu, Y. (2016). Lossless compression of public transit schedules. *IEEE Transactions on Intelligent Transportation Systems*, ***17***(11), 3075-3086.
6. Vallathan, G., & Jayanthi, K. (2015, December). Lossless compression based on hierarchical extrapolation for biomedical imaging applications. In *2015 International Conference on Microwave, Optical and Communication Engineering (ICMOCE)* (pp. 146-149). IEEE.
7. Moore, B. (1981). Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE transactions on automatic control*, ***26***(1), 17-32.
8. Jolliffe, I. T. (2002). *Principal component analysis for special types of data* (pp. 338-372). Springer New York.
9. Tzeng, J. (2013). Split-and-combine singular value decomposition for large-scale matrix. *Journal of Applied Mathematics*, *2013*.
10. Shi, H. (2009, August). Application of principal component analysis to general contracting risk assessment. In *2009 ISECS International Colloquium on Computing, Communication, Control, and Management* (Vol. 3, pp. 53-56). IEEE..

11. Tzimiropoulos, G., Zafeiriou, S., & Pantic, M. (2011, March). Principal component analysis of image gradient orientations for face recognition. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)* (pp. 553-558). IEEE.
12. Li, H. (2016). Accurate and efficient classification based on common principal components analysis for multivariate time series. *Neurocomputing*, ***171***, 744-753.
13. Kamencay, P., Trnovszky, T., Benco, M., Hudec, R., Sykora, P., & Satnik, A. (2016, May). Accurate wild animal recognition using PCA, LDA and LBPH. In *2016 ELEKTRO* (pp. 62-67). IEEE.
14. Santo, R. D. E. (2012). Principal Component Analysis applied to digital image compression. *Einstein (São Paulo)*, ***10***, 135-139.
15. Du, Q., & Fowler, J. E. (2008). Low-complexity principal component analysis for hyperspectral image compression. *The International Journal of High Performance Computing Applications*, ***22***(4), 438-448.
16. Lee, C., Youn, S., Jeong, T., Lee, E., & Serra-Sagristà, J. (2015). Hybrid compression of hyperspectral images based on PCA with pre-encoding discriminant information. *IEEE geoscience and remote sensing letters*, ***12***(7), 1491-1495.
17. Wang, C. W., & Jeng, J. H. (2012, November). Image compression using PCA with clustering. In *2012 International Symposium on Intelligent Signal Processing and Communications Systems* (pp. 458-462). IEEE.
18. Vaish, A., & Kumar, M. (2015, March). WDR coding based image compression technique using PCA. In *2015 International Conference on Signal Processing and Communication (ICSC)* (pp. 360-365). IEEE.
19. Lim, S. T., Yap, D. F. W., & Manap, N. A. (2014, September). Medical image compression using block-based PCA algorithm. In *2014 International Conference on Computer, Communications, and Control Technology (I4CT)* (pp. 171-175). IEEE.
20. Wu, M. S. (2014). Genetic algorithm based on discrete wavelet transformation for fractal image compression. *Journal of Visual Communication and Image Representation*, ***25***(8), 1835-1841.
21. Hussain, A. J., Al-Jumeily, D., Radi, N., & Lisboa, P. (2015). Hybrid neural network predictive-wavelet image compression system. *Neurocomputing*, ***151***, 975-984.
22. Prabhu, K. M. M., Sridhar, K., Mischi, M., & Bharath, H. N. (2013). 3-D warped discrete cosine transform for MRI image compression. *Biomedical signal processing and control*, *8*(1), 50-58.
23. Prabhu, K. M. M., Sridhar, K., Mischi, M., & Bharath, H. N. (2013). 3-D warped discrete cosine transform for MRI image compression. *Biomedical signal processing and control*, ***8***(1), 50-58.
24. Quijas, J., & Fuentes, O. (2014, April). Removing JPEG blocking artifacts using machine learning. In *2014 Southwest Symposium on Image Analysis and Interpretation* (pp. 77-80). IEEE.
25. Datta, D. (2014). Wavelet analysis based estimation of probability density function of wind data. *International Journal of Energy, Information and Communications*, ***5***(3), 23-34.
26. Zhou, Z., Hua, D., Wang, Y., Yan, Q., Li, S., Li, Y., & Wang, H. (2013). Improvement of the signal to noise ratio of Lidar echo signal based on wavelet de-noising technique. *Optics and Lasers in Engineering*, ***51***(8), 961-966.
27. Ibraheem, M. S., Ahmed, S. Z., Hachicha, K., Hochberg, S., & Garda, P. (2016, February). Medical images compression with clinical diagnostic quality using logarithmic DWT. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)* (pp. 402-405). IEEE.
28. Baviskar, A., Baviskar, J., Mulla, A., Jain, N., & Dave, P. (2016, March). Coded sub band replacement DWT based space image compression. In *2016 IEEE Aerospace Conference* (pp. 1-8). IEEE.