Dimension Reduction of Images Using Principal Component Analysis Algorithm

THET THET KHAING[1], PHYU SIN NYEIN[2], MYINT SOE KHAING[3], KHAING KHAING WAI[4]
[1]*ITSM Department, University of Computer Studies (Yangon), Myanmar*
[2]*Electronics Engineering Department, Yangon Technological University, Myanmar*
[3]*Faculty of Computer Science, University of Computer Studies (Yangon), Myanmar*
[4]*Faculty of Computer Systems and Technologies, University of Computer Studies (Yangon), Myanmar*

Abstract- ***Nowadays, the increasing volume of images is absolutely demanded in most of digital image processing and analysis. It is observed that high-resolution or high-dimensional images create many issues and challenges to deal with and, as a consequence, the compression of images has become an essential requirement in various applications related with images. In this paper, Principal Component Analysis (PCA), a dimensionality reduction algorithm, is applied in effectively compressing or reducing high-dimensionality of images. The main objective of the system is to show that applying PCA algorithm can efficiently perform not only in compressing of images but also in minimizing transmission time of images over the Internet. According to the experimental results, the transmission time of compressed images has achieved a significant improvement especially for the downloading activities via mobile devices.***

Indexed Terms- ***High-dimensional, High-resolution, Image Compression, and Principal Component Analysis.***

I. INTRODUCTION

Generally, images may be either raster or vector images or each image data is organized into two-dimensional matrix pixel values. As each pixel consists of its respective RGB bits value, high-resolution image can also be referred as high-dimensional data space. The main issue or challenge is that how to process high-resolution or high-dimensional representation of image data in predictive big data analytics because these images require greater bits for storage and transmission. A large number of computations and compressions of these images has become an essential requirement in various applications related with images. When images are increasing in today's multimedia age, the most significant visual features are at high-dimensional data space.

An image can typically be assumed as a two-dimensional matrix which can be specifically arranged in rows and columns. Every element of this matrix is called image element or pixel. Moreover, reducing redundant features from images become essential as processing all features directly can face a problem called "Curse of Dimensionality". It has also introduced many difficulties including computational complexity, feature sparsity and redundancy. To solve the problem, there is a need to use dimension reduction algorithms, a solution for high-dimensionality, which extract only essential features from the feature vectors of images. The results may offer the reduced number of features which best express the image which brings several advantages such as speeding up the computing algorithms' execution time [8].

Principal Component Analysis (PCA) can be applied as a dimensionality reduction algorithm preserving with as much as possible of the original information content. Moreover, PCA which used in reducing of dimensions allows the identification of standards in data and their expression. PCA has been successfully adopted in signal processing, image processing, system control theory, communication, pattern recognition, and so on [3].The paper is organized as follows: the related works which motivating the proposed work is briefly presented in section 2. The issues and challenges of high-dimensional data and how to reduce high-dimensionality of images are expressed in section 3. In section 4, background theories of principal component analysis (PCA) is

discussed. The explanations about experimentation of the system is then described in section 5. The final conclusion and future works are presented in section 6.

II. RELATED WORKS

High-dimensional or high-resolution image data has become a big challenge in transmission over Internet as large volumes of images are transferred every day. S.C Ng [5] evaluated PCA technique on high-resolution images' feature reduction performance comparing with feature-reduced images' visual quality in different variance values. When volume of images increases, Mohammed Amin Belarbi, Said Mahmoudi and Ghalem Belalem [2] expressed that dimension reduction method PCA can be applied to the reduction of SIFT and SURF image features ranging from 10% to 90% compression rate. In order to reduce the storage and get better image quality, Yaxiong Chen, Zhangcan HUANG and et.al. [10] Have proposed to apply the image compression algorithm combined with Contourlet Transform (CT). They discussed that the amount of information in the digital image is very large, it is important to apply the appropriate method to compress the image in order to store and transmit efficiently.Zebin Wu and et al. [11] explored that increasing volume of hyperspectral image data is absolutely required to consider in terms of storage and data processing. Thus, they suggested that large-scale hyperspectral image processing using PCAon parallel and distributed platforms to obtain better speedups when compared to traditional processing platform.

III. DIMENSION REDUCTION OF IMAGES

In current time, large scale high-dimensional or high-resolution images have encountered many difficulties to deal with as the higher resolution of the image is, the larger its data volume is. In fact, the number of features or dimensions of an image refer to the number of pixels composing of that image. It is also observed that images of larger sizes would require a large number of computations. As a consequence, compression of images has become an essential requirement in various applications related with images [2]. Images, in general, will consist of redundant data which means duplication of data in these images. It may be either repeating pixel across the image or pattern, which is more frequently repeated in image.

Reduction of redundant data in images help to achieve saving in storage space of that images. Image compression occurs when one or more of redundant pixels in image are reduced or eliminated. The quantity of pixels used in image representation can also be reduced without extremely change visualization of image. Therefore, dimension reduction or compression of images is typically an application of data compression which encodes the original image with reduced bits. Actually, an image is removed or group together its certain parts in order to reduce its size as sending and uploading an uncompressed image can take a long time indeed [5]. The amount of information in the digital image is very large, and it occupies a large amount of resources. So it is important to use appropriate method to compress the image in order to store and transmit. Image compression is divided into lossless compression and lossy compression. Principal Component Analysis (PCA) is a kind of feature vector based lossy compression algorithm utilized in many applications for image compression and digital watermarking. In recent years, PCA has attracted great attention in image compression. The main objective of image compression applying PCA is to reduce redundancy pixels of images for enhancing image sharing, transmitting and storing.

IV. PRINCIPAL COMPONENT ANALYSIS (PCA)

Typically, images can be represented as matrices and vectors composing a lot of pixel values. Principal Component Analysis (PCA) is applied to reduce the dimensionality of data matrix. It is a powerful tool for data analysis and pattern recognition and it can also provide as a statistical technique for image processing such as compression, dimension reduction and so on.PCA usually performs dimensionality reduction by extracting the principal components (PCs) of multi-dimensional data. PCA can also be efficiently applied as a digital image compression preserving with as much as possible of the original information content. According to PCA can represent high-dimensional data into a lower dimensional form, it can be applied to compress an image from original high-dimensional

representation (RGB values for each pixel) to lower dimensional representation [6].

The benefit of applying PCA is that it can perform lossless information not only finding patterns in data but also compressing data by reducing the number of dimensions [5]. PCA achieves compression through discarding the principle components with small eigenvalues. Generally, the steps which are followed in a PCA of a digital image are as follows:

- Correcting Image by the Mean
- Computing Covariance Matrix
- Calculating Eigenvalues and Eigenvectors
- Reconstructing Original Image

Algorithm: PCA for Image Compression

Step1: Obtain the feature column vector matrix from input image data.
Step 2: Obtain the covariance matrix A
Step3: Using characteristic equation$\lambda_i - EA = 0$to obtain the Eigen values. These Eigen values form $E_y$matrix.
Step 4: Considering the Eigen values to calculate the eigenvectors matrix.
Step 5: Obtain the TransformationB^T by considering the Eigenvectors as their columns.
Step 6: Obtain the features vector matrix by$C_v = C_A B^T$.
Step 7: For compression of an image, the dimensionality of the new feature vector is reduce by setting small Eigen values l to zeros.

V. EXPERIMENTATION

The original high-resolution image is shown in figure 1 is taken as an uncompressed input image and after applying step-by-step PCA procedures, the result image can be seen in figure 2. It can be observed that PCA actually reduces the less significant features or pixels and also remove redundant pixels from the image, however, principal features are still remained in it. In other words, the dimensions of the original image has been reduced without losing much information of that image. Moreover, the file size of image can be effectively compressed by PCA which can be compared in figure 1 and 2 without much loss of information. The file size of the dimension-reduced or compressed image has achieved 66.64% in reduction.

Fig.1The original input image (file size = 2500 KB)

Fig. 2The dimension-reduced or compressed image using PCA (file size = 1666 KB)

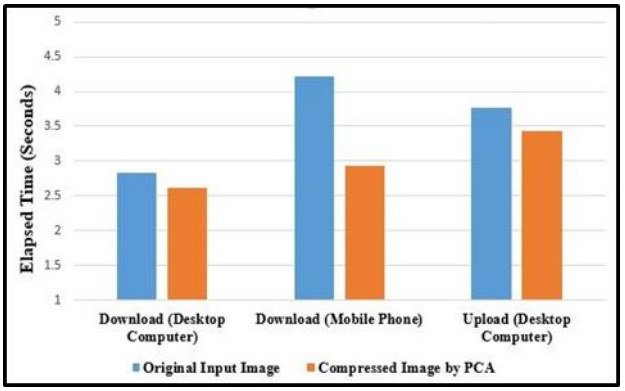

Fig. 3The comparison of elapsed time between uncompressed image and compressed image by PCA

The comparison results of the elapsed time for original uncompressed image and compressed image to upload and download over the Internet are shown in figure 3. In the figure, it can be clearly seen that the downloading time for original input image via desktop computer and mobile phone is longer than that of compressed image. Moreover, the uploading time for compressed image via desktop computer is faster than that of uncompressed image. According to the experimental results, it can be shown that the uploading time has improved 13.3% when the image has been compressed. The downloading time via desktop computer and mobile phone has decreased 9.29% and 28.9%respectively. Therefore, it can be proved that there has a significant improvement in elapsed time to download the compressed images by applying PCA for mobile users.

VI. CONCLUSION AND FUTURE WORKS

With the rapid development of Internet, the increasing volumes of images require a lot of storage space and transmission time. Moreover, it is observed that large scale high-dimensional or high-resolution images would require a large number of computations. In this system, Principal Component Analysis (PCA) algorithm is applied to efficiently compress large-scale images from original high-dimensional (RGB values for each pixel) to lower dimensional representation. According to the experimental results, applying PCA algorithm not only effectively reduce the file size of images without losing much information but also significantly improve in transmission time for compressed images. In future works, we will extend the system by experimenting the large-scale images from diverse sources by utilizing PCA on distributed platform to achieve better performance in compared with traditional processing platform.

REFERENCES

[1] B. K. Gunturk, et al., -Eigenface-based super-resolution for face recognition,| *in Proc. International Conference on Image Processing, IEEE*, 2002.

[2] B. Mohammed Amin, S. Mahmoudi, and B. Ghalem, -PCA as dimensionality reduction for large-scale image retrieval systems,| International *Journal of Ambient Computing and Intelligence (IJACI)*, vol. 8, pp. 45-58, 2017.

[3] G. Ajit P, S. R. Khot, -Facial expression recognition using principal component analysis,| *International Journal of Soft Computing and Engineering (IJSCE)*, vol. 3, pp. 2231-2307, 2013.

[4] N. A. Miranda, et al., -Image processing using Pearson's correlation coefficient: Applications on autonomous robotics,| *in Proc. 13th International Conference on Autonomous Robot Systems, IEEE*, April 2013, pp. 1-6.

[5] Ng. S. C, -Principal component analysis to reduce dimension on digital image,| *in Proc. Procedia computer science 111*, 2017, pp. 113-119.

[6] R. John A, -Covariance estimation in full-and reduced-dimensionality image classification,| *in Proc. Image and Vision Computing*, 2009, vol. 27, pp. 1062-1071.

[7] Santo, Rafael do Espírito, "Principal Component Analysis applied to digital image compression", 2012, pp. 135-139.

[8] Shereena, V. B, M. David Julie, -Significance of dimensionality reduction in image processing Signal & Image Processing,|*International journal (SIPIJ)*, vol. 6, June 2015.

[9] Y. Ming-Hsuan, et al., -Detecting faces in images: A survey,|*IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, pp. 34-58, 2002.

[10] Z. Huang, C. Yaxiong, et al., -Lossy image compression using PCA and contourlet transform,| *in Proc.MATEC Web of Conferences, EDP Sciences*, 2016, vol.54.

[11] Z. Wu, et al., -Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures,| *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, pp. 2270-2278, June 2016.

[12] www.kaggle.com

[13] www.unsplash.com

[14] www.peopleimages.com

[15] www.image-net.org