International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Image Compression Methods using Dimension Reduction and Classification through PCA and LDA: A Review

Khushboo Kumar Sahu¹, Prof. K. J. Satao²

¹Dept. of Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai 490024 C.G. India

²Prof. Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai 490024 C.G. India

Abstract: This paper presents in depth survey on various techniques of compression methods. Linear Discriminant analysis (LDA) is a method used in statistics, pattern recognition and machine learning to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be for dimensionality reduction before later classification. Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of variables of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The purpose of the review is to explore the possibility of a combined approach for image compression in which the best features of LDA and PCA shall be used. Another purpose of the study is to explore the possibility of image compression for multiple images.

Keywords: Image Compression, Dimension Reduction, Linear Discriminant Analysis (LDA), Principal Component Analysis (PCA)

1. Introduction

Image is a sequence of picture elements called pixels arranged generally in the form of rectangular matrix i.e. rows and columns. Each pixel represents the color intensity which forms the image as a whole.

In this article we focus on Image Processing. Image processing is processing of images on their physical and mathematical parameters by mathematical equations. The output of image processing may be either an image (modified in terms of physical or mathematical parameters) or a set of characteristics related to that image.

Digital image processing frequently applied in the modern applications rather than Analog image processing. In this digital era everything is computer oriented i.e. digitized. Mode of communication has been enhanced with digital equipment and network technologies. Much of the storage space is required to store large amount of digital data of an image. It is acceptable in the standalone storage system but for the communication taking place over various networks there is limited transfer capacity. Due to the problem of limited bandwidth there is a need of image processing before it is transferred. To overcome these kinds of situations several techniques have been developed in image processing.

In recent years, digital images and videos have gained more popularity in social media, data mining, detection, and in networks. Different types of image editing software had also gained importance. That is why Image compression has become a necessity due to the increasing demand on data transfer and storage. In image processing there is one specific field by which size of data can be reduced called as Image Compression. These methods use various mathematical models in order to reduce irrelevance and redundancy of image data, so that it can be stored or transmitted efficiently.

Image compression can be of the following kind, lossy or lossless. Lossless compression is generally preferred for backup storage, warehousing, and archival purposes e.g. medical imaging, technical drawings, clip art, or comics. Whereas Lossy methods are especially used for natural images e.g. personal digital images, wallpapers, etc. photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate.

There may be different types of redundant data in an image. With the application of compression techniques they can be reduced, types of redundant data are as follows:

- 1. Coding Redundancy
- 2. Inter Pixel Redundancy
- 3. Psycho Visual Redundancy

Coding Redundancy is a type of redundancy in which image data is encoded in such a manner that resultant bits are less than the actual image bits.

Inter-pixel Redundancy also known as spatial redundancy, inter frame redundancy, or geometric redundancy – the intensity of a pixel may be strongly correlated to its neighbor's intensity value. In this method we try to predict the intensity value of any given pixel by its neighbor. So we need not to store the absolute intensity values rather we can use changes present in the intensity values.

Psycho-visual Redundancy as we know that the human eye does not respond to all visual information (color saturation etc.) with equal sensitivity. Means that eye is less sensitive to the higher frequencies and more to the lower frequencies. So it can be reduced without making any significant difference to the human eye.

2. Image Compression

As we discussed above that Image compression techniques can be of two types:

- 1. Lossless-Compression Method
- 2. Lossy-Compression Method

Lossless Compression Method

When the image data is encoded in such a manner that it does not lose its quality i.e. image can be restored from the reduced data after decompression. The various methods which can be used as lossless compression modes are:

- **1. Variable Length Coding:** In this method different symbol of image is encoded with the variable length code words. While shorter code words are assigned to the most frequent symbols.
- **2. Run Length Encoding:** In this method image symbols are replaced by a sequence (run) of identical symbols which are attributed with pair of values containing the symbol and the run length (i.e. count), Used in images containing homogeneous regions.
- **3. Differential Coding:** It explores the inter pixel redundancy in digital images. Here we apply difference operator to neighboring pixels at any pixel position to calculate a difference image.
- **4. Predictive Coding:** It also explores the inter pixel redundancy in digital images. Here the basic idea is that we encode only the new information in each pixel position rather than storing the complete information. The difference between the actual and the predicted value of the pixel is considered to be the new information.

Lossy Compression Method

When some amount of deterioration in the image visual quality is acceptable then lossy compression methods are used. Here the actual image cannot be reformed after decompression. Lossy compression techniques deliberately introduce a certain amount of distortion to the encoded image (generally tries to explore the psycho-visual redundancy in the original image). But There must be an appropriate relation between the amount of error (loss) and the resulting bit savings otherwise the objective cannot be achieved. The different types of lossy compression methods are:

- **1. Quantization:** The input data is partitioning into range of smaller set of values. There are two types of this Quantization method: Scalar and Vector.
- **2. Transform Coding:** It works directly on the pixel values of image data. It uses a reversible-linear mathematical transform to map the pixel values onto a set of

coefficients. After the computation of coefficients values the resultant are then quantized and encoded.

- **3. Fractal Coding:** Different attributes of an image such as color separation, edge detection and texture are used to decompose the image into segments.
- **4. Wavelet Coding:** In Wavelet coding the coefficients of a transform that de-correlates the pixels of an image are coded more efficiently than the original pixels.

3. Dimensionality Reduction

In statistics Dimensionality Reduction or Dimension Reduction is the process of minimizing or reducing the number of random variables under consideration which is further divided into feature selection and feature extraction.

In statistics feature selection also known as variable selection, attribute selection or variable subset selection, is the process of selecting a subset of relevant features (variables, predictors) for use in model construction. Feature selection techniques are used for simplification of models to make them easier to interpret by researchers/users, shorter training times, enhanced generalization by reducing over fitting (formally, reduction of variance).

In pattern recognition/image processing, feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant. When input data for a program is too large to be processed and it is suspected to be redundant (e.g. the repetitiveness of images presented as pixels), then it can be transformed into a reduced set of features (also named a features vector). This process is called feature extraction. The extracted features are expected to contain the relevant information from the input data, so that the desired task can be performed by using this reduced representation instead of the complete initial data.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is an unsupervised dimensionality reduction method, means it does not require classes/variables to be known in advance. PCA seeks to reduce the dimension of the data by finding a few orthogonal linear combinations (the principal components PCs) of the original variables with the largest variance. As per the number of the original variables there are as many PCs. The first several PCs explain most of the variance, so that disregarded the rest can be with minimal loss of information, for many datasets to reduce the dimensionality of the huge data along with retaining as much information as possible in the original dataset.

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a linear projection of multi dimensional space to a low dimensional space. It is a supervised technique for classification, where we try to maximize the class scatter and minimize the within-class scatter.

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Research works had been carried out by several authors for improving image compression using PCA, LDA, 2D-PCA for gray scale and colored images. Md. Mofarreh [3] [10] and Telgaonkar Archana H. [12] has introduced some PCA and LDA algorithm to compress color images. High compression ratios can be obtained by using PCA method. Unsupervised PCA and Supervised LDA methods can be used to dimension reduction whereas accuracy of predicted values of classifier is analyzed on specific factor given.

A.Dwivedi et al. (2006) observed that Multi dimensional Principal Component works directly on the vector data of color image where each band is taken as dimension of the matrix. The work is done on the principle of applying 2DPCA and the methodology is tested on several standard images (lena, mandril, etc.). In the observation it is found that the quality of reconstructed image is good with more numbers of eigenvectors but compression ratio is compromised. Overall the method was better than standard PCA technique and the other performance measures (i.e. computational time, compression ratio) are also improved [2].

Md. Mofarreh et al. (2012) found that the speed of algorithm can be increased with parallel programming, because the compression process of the various band of the image are independent with each other. The mentioned method can be applied for images and a new format for images can be produced and short time of reconstruction of compressed images in comparison with the other formats such as JPEG [3].

Chiyuan Zhang et al. (2013) found that TEM (Total Error Minimization) algorithm and its improved version TEM-Compensate (TEM-C) for image compression has the key advantage over previous methods comes from the maximum exploitation of the full label set (i.e., the colors for all the pixels) at the encoding stage. Furthermore, TEM-C used the label set to generate and store a difference image for correcting the prediction error and improved the colorization quality significantly. Experimental results demonstrated the outstanding performance of the proposed methods. Although the computation burden is still high, TEM-C is already competitive to the industrial standard JPEG in image quality and compression ratio [4].

Ms.Pallavi et al. (2013) discussed about the image classification, wavelet compression and convert an image into an array using Delphi image control tool. They proposed a methodology which is created in Delphi to implement Huffman coding. In the experiment they found that Huffman coding of compression technique is the best technique in lossless compression and the process is complete in two passes to compress any file. The Wavelet Compression technique was used in this study [5].

Maryam Imani et al. (2014) observed that the proposed method that is called principal component discriminant analysis (PCDA) improves the classification accuracy and works better than both PCA and LDA. The experimental results obtained by using two hyper spectral data (an urban image and an agriculture image) shows the good efficiency of proposed method, here the principal components of PCA i.e. components with smaller variance that have useful classification information can be used in Discriminant Analysis to lower power components of PCA Called as Principal Component Discriminant Analysis (PCDA). Using the following method the classification accuracy can be improved by the lower power components [9].

Md. Mofarreh et al. (2015) found that High compression ratios can be obtained by using PCA method. A new PCA based (Extended PCA) method to compress color image, this can be utilized in parallel mode to increase the compression speed.

More precise method for selecting the bands of image can be utilized to improve the performance of compression.

Extended PCA based method to compress color images which can be utilized to compress single image rather than a set of separated images. This method uses the correlations between three color components of an image. This method can be utilized in parallel mode to increase the compression speed [10].

Jagruti Rajput et al. (2015) found that in the proposed technique a combination of image registration methods PCA and GPOF for Multi frame super-resolution one can measure the translation parameter and allow large pixel motion while keeping the image neighborhood relatively small. After that PCA can be used as compression method. The Image Fusion is used to get one output image with multiple low resolution images. Image interpolation using Bi-cubic method is used to get high resolution or reconstruct the image [11].

Telgaonkar Archana H et al. (2015) found that Unsupervised PCA and Supervised LDA methods can be used to dimension reduction. By the use of stated methodology performance analysis has been done on high dimensional image dataset, also the accuracy of predicted value of classifier is measured based on specific factor [12].

4. Discussion

We observed that PCA and LDA can be applied to image, color image in many different ways to get the better classification of image and in order to achieve better compression performance. Experiment has been done only for single image i.e. gray scale (one channel), and color image (multi channel). Various methods have been proposed for the image compression but still there is no study has been done to compress multiple images, large amount of redundant data can be found in the form of pixel values from multiple images which can be reduced in order to get maximum compression ratio.

5. Conclusion

An image can be pre-processed before it can be used further such as storage, transmission, etc. PCA as a statistical tool can increase the efficiency of image compression while LDA can be used as classifier. Using PCA and LDA we can improve the performance of various compression techniques as a combined approach.

References

- L. Wang, X. Wang, and J. Feng. On Image Matrix Based Feature Extraction Algorithm. IEEE Transactions On Systems, Man, And Cybernetics part B: Cybernetics, Vol. 36, No. 1, Feb-2006.
- [2] A. Dwivedi, et. Al. "Color Image Compression Using 2- Dimensional Principal Component Analysis (2DPCA)" Proc. of ASID, pp.488-491, October 2006.
- [3] Md. Mofarreh. A new technique for Image Compression using PCA. IJCSCN, ISSN: 2249-5789, Volume 2 (1), P.P. 111-116, Feb-2012.
- [4] Chiyuan Zhang, Xiaofei He. Image Compression by Learning to Minimize the total Error. IEEE, Vol.23, No. 4, Apr-2013.
- [5] Ms. Pallavi M. Sune Prof. Vijaya K. Shandilya "Image Compression Techniques based On Wavelet and Huffman Coding" International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 4, April 2013.
- [6] R. Seth, S Shantaiya, Survey on Image Compression Methods with PCA & LDA. IJSR Vol.4 Issue 4 April 2015
- [7] K. Sau, R. K. Basak, A. Chanda. Image Compression Based on Block Truncation Coding using Clifford Algebra. Elsevier, Procedia Tech 10 (2013) 699-706.
- [8] H. Yue, X. Sun, J. Yang et. al.. Cloud-Based Image Coding for Mobile Devices. IEEE, Vol.15, No.4, Jun-2013.
- [9] Maryam Imani, Hassan Ghassemian. Principal Component Discriminant Analysis for Feature Extraction and Classification of Hyperspectral Images. IEEE, 978-1-4799-3351, ACC. No. -14253352 Feb-2014.
- [10] Md. Mofarreh. Color Image Compression using PCA.IJCA (0975-8887), Volume 111- No. 5, P.P 16-19, Feb-2015.
- [11] Jagruti Rajput. Image Registration using combination of PCA and GPOF method for Multiframe Super Resolution. IJCA (0975-8887), Volume 120 - No. 12, P.P. 01-05, Jun-2015.
- [12] Telgaonkar Archana H. Dimensionality Reduction and Classification through PCA and LDA. IJCA (0975-8887), Volume 122 - No.17, P.P. 04-08, Jul-2015.
- [13] N. Pati, A. Pradhan, L.K. Kanoje, T Das. An approach to Image Compression by using Sparse Approximation Technique. Elsevier, Procedia Comp. Sci. 48 (2015) 769-775.
- [14] A.G. Nath, M. S. Nair, J. Rajan. Single Image Super Resolution from Compressive Sample using Two Level Sparsity based Reconstruction. Elsevier, Procedia Comp. Sci. 46 (2015) 1643-1652