Importance of Dimensionality Reduction in Image Processing

Jigna J. Patel
Asst.Prof.
Dr. S &S.S.Ghandhy GEC, Surat
jigna2012me@gmail.com

Abstract: This paper presents a 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 classifies an object into two or more classes. This results in a dimensionality reduction before later classification.Principal component analysis (PCA) uses an orthogonal transformation to convert a set of 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 image compression for multiple images.

Keywords: *Principal Component Analysis,Linear Discriminant Analysis,Image Compression,Redundancy,lossy compression,lossless compression*

*****

I. Introduction:

An image is formed by several pixels i.e. picture elements that are arranged sequentially in the form of rectangular matrix i.e. rows and columns. These pixels represent the color intensity of the image. In Image processing, the processing of images is carried out on their physical and mathematical parameters by mathematical equations. We may get as an output of image processing either an enhanced image or a set of characteristics related to that image. Nowadays, the mode of communication is digital therefore much storage space is required to store large amount of digital data of an image.Before transferring the image, image compression is required because of the problem of limited bandwidth .Several techniques have been developed in image processingto overcome these kinds of situations. As image editing softwares are easily available in the market, image compression has become a necessity for data transfer and storage.

To transmit data efficiently, image compression can be applied to images in image processing. Image compression uses use various mathematical models in order to reduce irrelevance and redundancy of image data.

Image can be compressed in two ways using,

i) Lossless compression: lossless compression is mostly required for text and data files, such as bank records and text articles .e.g. medical imaging, technical drawings, clip art, or comics.

ii) Lossy compression: Lossy compression is acceptable in many imaging application such as audio, video, and images. In this type of compression imperceptible loss of fidelity is acceptable to achieve a substantial reduction in bit rate.

Different types of redundancy can be found in data in an image. By applying the compression techniques they can be reduced, types of redundant data are as follows:

1. Coding Redundancy: Coding redundancy is associated with the representation of information i.e. if the gray levels of an image are coded in a way that uses more code symbols than needed to represent each gray level then the resulting image is said to contain coding redundancy. Here, the image data is encoded in such a manner that resultant bits are less than the actual image bits.

2. Inter Pixel Redundancy: Inter-pixel Redundancy is also called spatial redundancy, inter frame redundancy, or geometric redundancy where the intensity of a pixel may be greatly correlated to its neighbor's intensity value. We can predict the intensity value of any given pixel by its neighbor in inter-pixel redundancy. So, we don't need to store the absolute intensity values, instead we can use changes present in the intensity values.

3. Psycho Visual Redundancy: Psychovisual redundancies exist due to the human perception as human eye does not respond to all visual information with equal sensitivity. This means, eye is less sensitive to the higher frequencies and more to the lower frequencies. Therefore, it can be reduced without making any significant difference to the human eye.

II. Image Compression

We have already discussed two types of Image compression techniques as follows:

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

Lossless Compression Method –In this method, the image data is encoded in such a manner that it does not lose its quality. Original image can be restored from the reduced data after decompression.

Following are the methods that can be used as lossless compression modes are:

1. **Variable Length Coding:** Variable length code words are used for encoding different symbol of image. In this method, shorter code words are assigned to the most frequent symbols.
2. **Run Length Encoding: RLE** is a simple form of lossless compression technique where runs of data (i.e. sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value. It is most useful on data that contains many such runs but not useful with files that don't have many runs as it could greatly increase the file size.
3. **Differential Coding:** It explores the inter pixel redundancy in images where data is represented by storing the only the first value in the data stream, all other values are represented by the difference between the value and the preceding value.
4. **Predictive Coding:** It also explores the inter pixel redundancy of near located pixels in digital images. In this type of coding only the new information included in the new pixel is extracted and stored.The difference of the pixel and the predicted value of that pixelwill be used to define this information. Here, the prediction is based on the known values of the closely located pixels.

Lossy Compression Method -lossy compression methods can be used where some amount of deterioration in the image visual quality is acceptable because here the actual image cannot be reformed after decompression. Lossy compression techniques introduces a certain amount of distortion to the encoded image but there must be an appropriate relation between the amount of error (loss) and the resulting bit savings otherwise the purpose cannot be achieved.

The different types of lossy compression methods are:

1. Quantization: Quantization is the process of mapping a continuous or discrete scalar or vector, produced by a source, into a range of smaller set of values. Scalar quantization and Vectorquantization are the two types of quantization methods.

2. Transform Coding: It works on the pixel values of image data where it uses a reversible-linear mathematical transform to map the pixel values onto a set of coefficients. The resultant value obtained after the computation of coefficients values are then quantized and encoded.

3. Fractal Coding: This lossy compression method used for digital images is based on fractals.Different attributes of an image e.g. color separation; edge detection and texture are used to decompose the image into segments.

4. Wavelet Coding: Here, the coefficients of a transform that de-correlates the pixels of an image are coded more efficiently than the original pixels.

III. Dimensionality Reduction

The process of minimizing or reducing the number of random variables under consideration by obtaining a set of principal variables is known as Dimensionality Reduction or Dimension Reduction in statistics and machine learning. It is further divided into feature selection and feature extraction. In statistics feature selection is the process of selecting a subset of relevant features or variables for use in model construction. It is also known as variable selection, attribute selection or variable subset selection. Feature selection is useful for simplification of models so that the researchers can easily interpret them. Feature extraction starts from an initial set of measured data in machine learning, pattern recognition and in image processing. It than builds derived values or features that is intended to be informative and non-redundant. This leads to subsequent learning and generalization steps which help in better human interpretations. When we have to process a large amount of input data which is redundant, then it can be transformed into a reduced set of features. This transformation process is called feature extraction. The extracted features contain 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 Components Analysis (PCA) is accomplished using mathematical formulation used in the reduction of data dimensions. Using PCA formulation the digital image may be compressed with a low level of loss. Such a reduction is advantageous for image compression, data representation; calculation reduction necessary in subsequent processing, etc.PCA is an unsupervised dimensionality reduction method as it does not require classes or variables to be known in advance.PCA reduces the dimension of the data by finding a few orthogonal linear combinations i.e. the principal components PCs of the original variables with the largest variance. The first several Principal Component's explain most of the variance, so that even if the rest are of minimal loss of information, for many datasets to reduce the dimensionality of the huge data and

retaining as much information as possible in the original dataset.

Linear discriminant analysis (LDA) is a supervised technique for classification , 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 resultant may be used as a linear classifier, or for dimensionality reduction before later classification.

IV. Related Work:

In 2013, Steven Fernandes and Josemin[2] had reported performance analysis of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for face recognition. They had carried out the analysis on various current PCA and LDA based face recognition algorithms using standard public databases. They authors used ORL and SHEFFIELD database consisting of 100 components and find it the best. Illumination Adaptive Linear
Discriminant Analysis (IALDA) gave them the best face recognition rate of 98.9% on CMU PIE database, and they got 98.125% recognition rate using Fuzzy Fisherface through genetic algorithm on ORL database.By the experiments carried out by the authors they find out that the best algorithm to recognize image without disturbance ,but to recognize image with disturbances, LDA is better and in terms of time taken, PCA tends to be much better than LDA, especially to recognize images with background disturbance .

In 2013, Ms. Pallavi M. Sune Prof. Vijaya K. Shandilya[3] discussed about the image classification, wavelet compression and convert an image into an array using Delphi image control tool. They had proposed a method created in Delphi to implement Huffman coding. In their experimental analysis, 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.

In 2014, Hang Su and Xuansheng Wang [4], proposed a new PCA method based on the linear discriminant analysis (LDA) space. They further derived from their theoretic analysis and numerical experiments that their new PCA method (PCA-LDA) can worked more effectively and efficiently. In PCA-LDA algorithm that they used, the principal components were obtained from the LDA space. Compared with original PCA, the PCA-LDA they got higher recognition rates.

In 2015, Md. Mofarreh et al.[5] found that by using PCA method ,High compression ratios can be obtained. Authors proposed a new PCA based (Extended PCA) method to compress color image, which can be utilized in parallel mode to increase the compression speed. They also found that more precise method for selecting the bands of image can be utilized to improve the performance of compression.

In 2015, Shereena V. B. and Julie M. David[6] presented a comparative study of dimension reduction methods namely Principal Component Analysis and Linear Discriminant Analysis. The authors had computed eigenvectors from their dataset and collected them in scatter-matrices, the between-class scatter matrix and within-class scatter matrix in LDA. They had trained the network about the reduced feature set (using PCA or LDA) of all the images from the training dataset in Neural Network based PCA/LDA. They had used Levenberg-Marquardt back propagation training algorithm with 3 hidden layers. From the experimental analysis done by them, they found that the rates are high for all the cases of PCA when compared to the LDA.

V. Discussion

From the study of PCA and LDA, it can be found that both PCA and LDA can be applied to images in many different ways to get the better classification of image and in order to achieve better compression performance. Until now many 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.

VI. Conclusion

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

References:

[1] Khushboo Kumar Sahu, Prof. K. J. Satao," Image Compression Methods using Dimension Reduction and Classification through PCA and LDA: A Review", IJSR'2013.

[2] Steven Fernandes and Josemin Bala, "Performance Analysis of PCA-based and LDA based Algorithms for Face Recognition", IJSPS'2013.

[3] Ms. Pallavi M. Sune Prof. Vijaya K. Shandilya,"Image Compression Techniques based On Wavelet and Huffman Coding ", IJARCSSE'2013.

[4] Hang Su and Xuansheng Wang,"Principal Component Analysis in Linear Discriminant Analysis Space for Face Recognition", DOI 10.1109/ICDH.2014.13 IEEE'2014.

[5] Mostafa Mofarreh," Color Image Compression using PCA", International Journal of Computer Applications (0975 – 8887) Volume 111 – No 5' February 2015.

[6] Shereena V. B.and Julie M. David,"Significance of dimensionality reduction in image processing", SIPIJ'2015.

[7] Mohammad Mofarreh Bonab and Mostafa Mofarreh-Bonab," A New Technique for Image Compression Using PCA, "International Journal of Computer Science & Communication Networks,Vol 2(1), 111-116.

[8] Amritpal Kaur, Sarabjit Singh and Taqdir," Face Recognition Using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) Techniques", IJARCCE' 2015.

[9] Chiyuan Zhang, Xiaofei He," Image Compression by Learning to Minimize the total Error",IEEE, Vol.23, No. 4'2013.

[10] Telgaonkar Archana H.," Dimensionality Reduction and Classification through PCA and LDA",IJCA (0975- 8887), Volume 122 - No.17, P.P. 04-08'2015.