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

Optimization of Huffman Encoding for Multiple Images Compression using PCDA

Khushboo Kumar Sahu[1], K. J. Satao[2]

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

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

Abstract: ***There are different classification and methods of image compression techniques. By using these compression methods we can reduce the redundant data from the file while preserving the same amount of information as it was before the compression. By these methods we can save the lots of storage space required to store the file and also the bandwidth required to send the data over the network. Principal component analysis (PCA) is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. 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. In this paper we are using the combination of PCA & LDA as PCDA approach where we compress the multiple images such that the compression ratio is greater than before and time requirement is lesser than it is for these individual images. More compression ratio can be obtained while working with multiple files rather than unitary files.***

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

1. Introduction

Every digital data is made up of character, symbols, text, diagrams and images, each having their own significance according to the need of the information in relevant area. An Image data is a different than all of them because they come with lots of variety such as different image quality color, type, size, dimension, channel, etc. The image is stored in the form of pixel (i.e. the picture element); each pixel is represented as some value of corresponding color value which makes the picture as a whole.

In this paper we focus on Image Processing. Image processing algorithms usually works on the physical and mathematical characteristics of an image using some form of mathematical equations. The resultant of the image processing algorithm process may either be a new image (modified in terms of physical or mathematical parameters) or a new set of values related to the image.

As we know that even a smallest picture we use in any field may have some amount of redundant information in terms of pixel value which can be better encoded such that we can reduce the size of the picture without even losing its quality. This kind of technique is known as Image Compression Technique. Also there are some applications or requirements where we can allow some amount of loss of quality in picture just to get more reduction of storage bits, this type of Image Compression method is known as Lossy if no information is lost (i.e. quality) then it is call Lossless Compression.

For applications such as archival storage, warehousing, digital library one can use lossless image methods and where some minor Loss is allowable such as natural images e.g. personal images, wallpapers, photographs then one can use lossy methods of image compression. There are mainly three different types of redundant information for an image file. And by applying the different image compression technique these redundancies can be reduced:

1) Coding Redundancy,
2) Inter Pixel Redundancy,
3) Psycho Visual Redundancy.

Coding Redundancy in this type of redundancy the algorithm encoded the image in such a way that resultant bits is lower than that of actual image bits.

Inter-pixel Redundancy in this type of redundancy a pixel's intensity value is correlated to its surrounding pixels; Leading us farther to find a common region to map those values in an intelligent way. Also we can make use of prediction to predict the value of any pixel by its surroundings.

Psycho-visual Redundancy in this type of redundancy the psycho visual parameters such as higher frequencies, lower frequencies, color saturation etc. are considered. So that it can be reduced without affecting the significant information of the image.

2. Dimensionality Reduction through PCDA

PCA and LDA can be used in combination, called PCDA, now we discuss here its working mechanism and how it's going to be used in combined form and also what purpose does it fulfill.

- **PCA:** "**Principal component analysis"** is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set.
- **LDA**: "**Linear Discriminant Analysis"** utilizes the label information in finding informative projection. It is used to find a linear combination of features which characterizes or separates two or more classes of objects or events.

In this combined approach, we apply both the algorithm together one after the other and make the better statistical tool.

A. Principal Component Analysis (PCA)

PCA "Principal component analysis" is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set". The PCA algorithm identifies the pattern in data which helps to express the data in a form where one can observe the similarities and variations among data set. And when we find such pattern and relation among the data then we can use those pattern and relation to reduce the data set either by reducing/ changing the dimension or either by selecting the less value representing the overall variability.

Algorithm Steps

Step1: Take multiple Images as an Input
Step2: combined them into one single file having the squared dimension
Step3: Calculate the ***Mean*** for the file
Step3: Calculate the ***Standard Deviation (StDev)*** of the file
Step4: Calculate the covariance matrix, ***CoMat***
Temp = (Mean-StDev)*Transpose(Mean-StDev);
CoMat = (Temp)/(N-1); where N= New Row/Column Size of Clubbed file
Step5: Then Calculate ***Eigenvector*** of the Resultant Matrix (***Co Mat***)
Step6: Calculate ***Principal Component (PC)*** for the entire file
PC = (Mean-StDev) * Row Value of the Eigen Matrix

B. Linear Discriminant Analysis (LDA)

LDA can be used to reduce dimensionality of the data while it tries to preserve more of classes' discriminatory information. LDA & PCA are closely related to each other but seeks for different pattern in data, although they both look for linear combinations of random variables which can best explain the data. Here we try to make a linear combination of predictors which can best separate our targets, while PCA does not take any account for differences in targets.

Algorithm Steps

Step 1: Take the input file created from above PCA steps
Step 2: Find the **newMean Vectors** of the resultant matrix
Step 3: Average of the **newMean** for the vectors
Step 4: Calculate the **newVariance**
Step 5: Calculate the **eigenvector** & corresponding Eigenvalues
Step 6: Sort the Eigenvalues with decreasing value of Eigenvectors
Step 7: Now form the new subspace by using this new Eigenvector

3. Huffman Encoding

Huffman encoding is an entropy encoding mechanism which is one of the methods of lossless data compression techniques. Huffman encoding technique is best suited for data as well as image data for compression. This method is based on the frequency of data item presents in data set. This method gives a new bit pattern to represent the data item available in data set, it tries to assign new bit pattern but lower in size which occurs more frequently than those which are less frequent, in this way we save more of the length to store the available information. This method creates a new database for available data known as code dictionary, where each data symbol is associated with a new code word assigned by the algorithm. Also there is uniqueness of code word assigned to each of the symbol present in the data set.

In the Huffman encoding technique when the first phase takes place to process the raw data and tries to build the statistical model the process becomes slow down. This is because of size of the raw data; we know that there is always a huge amount of redundant data. So the efficiency & accuracy phase one gets affected by it and phase two is totally dependent on phase one the statistical model.

Our sole aim in this study is to rectify the process of phase one where the statistical model for the raw data is made and to perform the optimization for this phase. So the Huffman encoding performance can increase and also the accuracy of the statistical model of phase one. Huffman encoding is very efficient & easy to implement, lossless technique of compression and most popularly used. Yet there are some problems like the process becomes relatively slow or overhead problem of the Huffman tree (if more unique data are present).

Algorithm Steps:

Step 1: Find the each symbol and its probability to be encoded present in our data.
Step 2: Now arrange each symbol in descending order of their probabilities and make them terminal nodes.
Step 3: Replace two symbols having smallest probabilities into a subgroup.
Step 4: Now assign value zero to top and value one to bottom branch.
Step 5: Repeat step 4 & 5until there no more node to merge.
Step 6: After the completion of the tree read branches from root to leaf to generate codeword's.
Step 7: Stop.

4. Result

We are using multiple images to club them into a single file and after that we are applying the PCDA process which in regards to those individual images gives better result. The result of single and multiple images have been shown below.

Figure 1: Multiple Clubbed Images

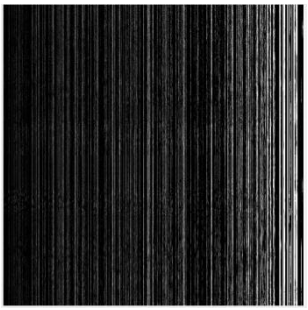

Figure 2: PCA Result of Clubbed File

Figure 3: After Compression of Clubbed File

Here we have used multiple squared images of same size, clubbed together and then we have applied all the three steps of proposed methodology PCA, LDA and Huffman encoding.

The compression ratio is to measure the outcome of the encoding technique, where we see the difference made by the proposed algorithm. To measure the performance of the data compression algorithm we just compare the number of bits required to represent the data before and after compression takes place. One can measure the performance of compression algorithm by checking the value of compression ratio, that how much compression on bits has been done.

The Compression Ratios Are:

No.	Image(s)	Compression Ratio	
		Huffman	With PCDA
1.	Cameraman	1.20	1.72
2.	Baboon, Lena, Cameraman, Iris	1.83	2.48

5. Discussion

The research work shows improvement of Huffman encoding technique with its statistical model by reducing the dimensionality of raw data it also shows that we can get more of compression ratio while we work on multiple images set rather than individual images.

6. Conclusion & Future Work

The research work which is carried out is concerned with the improvement of data preprocessing and dimension reduction, more specifically this study provides a novel idea to improve the performance and development of Huffman encoding.

The raw data often comes with redundant data but not surely all the time, and if we just go with the complete set of those raw data then it may be not worthwhile. So in order to improve the efficiency and accuracy of the encoding we can perform some extra steps by which we can reduce the complexity, by removing redundant data which is not necessary to represent the overall information available. The new idea proposed a statistical tool which is applied on the input data collected from the multiple image files and by performing that process over multiple files gives us more benefits than performing on one to one. This study has been done only on gray scale images; new algorithm can be proposed to applying on color images which has multiple bands. This work can be further extended to different file formats of images as well as different sizes of the file i.e. aspect ratios.

References

[1] 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] R. Seth, S Shantaiya, Optimization of Image Compression Technique: Huffman Coding by Using Pcda. IJSR Vol.4 Issue 3 March 2015
[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.
[15] K. Sau, R. K. Basak, A. Chanda. Image Compression Based on Block Truncation Coding using Clifford Algebra. Elsevier, Procedia Tech 10(2013) 699-706.
[16] Khushboo Kumar Sahu, Prof. K.J. Satao, Image Compression Methods Using Dimension Reduction and Classification through PCA and LDA : A Review. IJSR Vol.5 Issue 5,2016