A Survey and Study of Image Compression Methods

K.N. Abdul Kader Nihal[1], Dr. A.R. Mohamed Shanavas [2]

[1](Assistant Professor of Computer Science / Jamal Mohamed College, Trichy- 620020)
[2] (Associate Professor of Computer Science / Jamal Mohamed College, Trichy- 620020)

Abstract*: This paper focuses on the area of Image Compression which is used in various Image Processing Applications. This paper presents a survey and study of various image compression techniques primarily, The Principal Component Analysis approach. The PCA approach can be implemented in two ways namely, PCA Statistical Approach and PCA Neural Network Approach. It also deals with various benefits of using image compression methods.*

Keywords: *Image Compression, Principal Component Analysis Approach, PCA Statistical Approach, PCA Neural Network Approach*

I. Introduction

1.1 Image

Image can be defined as a two dimensional signal processed by the human visual system. The signals which are in analog form need to be converted to digital form in order to process, store and transmit. A digital image is a two dimensional array of pixels and each pixel is a digital value..

Data may be represented as an image which conveys more meaning. Visualizing objects is more suitable and easy to observe, easy to access, searchable hence the need to discover new methods for efficient storage and quick transmission of image data.

Images from the significant part of data, particularly in remote sensing, biomedical and video conferencing applications. The use of and dependence on information and computers continue to grow, so too does our need for efficient ways of storing and transmitting large amounts of data

1.2 Image Compression

Image compression addresses the problem of reducing the amount of data required to represent a digital image. [1] It is a process intended to yield a compact representation of an image, thereby reducing the image storage/transmission requirements. Compression is achieved by the removal of one or more of the three basic type of data redundancies:

1. Coding Redundancy
2. Interpixel Redundancy
3. Psychovisual Redundancy

Coding redundancy is present when less than optimal code words are used. Interpixel redundancy results from correlations between the pixels of an image. Psychovisual redundancy is due to data that is ignored by the human visual system (i.e. visually superfluous information).

Image compression methods reduce the number of bits required to represent an image by taking advantage of these redundancies.

An inverse process called decompression (decoding) is applied to the compressed data to get back the reconstructed image. The objective of compression is to reduce the number of bits as much as possible, while keeping the resolution and the visual quality of reconstructed image as close to the original image as possible. Image compression systems are composed of two distinct structural blocks: an encoder and a decoder.
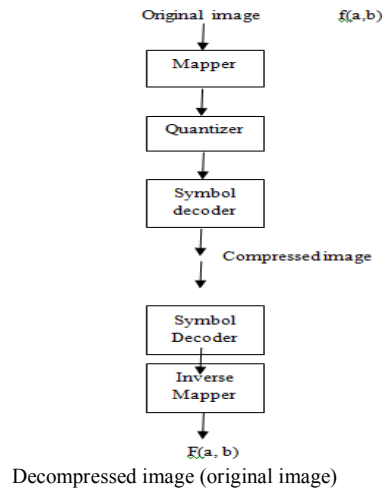
```
Original image          f(a,b)
        |
        v
   +-----------+
   |  Mapper   |
   +-----------+
        |
        v
   +-----------+
   | Quantizer |
   +-----------+
        |
        v
   +-----------+
   |  Symbol   |
   |  decoder  |
   +-----------+
        |
        v          Compressed image
        |
        v
   +-----------+
   |  Symbol   |
   |  Decoder  |
   +-----------+
        |
        v
   +-----------+
   |  Inverse  |
   |  Mapper   |
   +-----------+
        |
        v
     F(a, b)
```

Decompressed image (original image)

Image f(a,b) is fed into the encoder, which creates a set of symbols from the input data and uses them to represent the image. If we let n1 and n2 denote the number of information carrying units (usually bits) in the original and encoded images respectively, the compression that is achieved can be quantified numerically via the Compression Ratio,

CR = n1 /n2

As shown in the figure, the encoder is responsible for reducing the coding, interpixel and psychovisual redundancies of input image. In first stage, the mapper transforms the input image into a format

1.3 Benefits of compression
- It preovids a potential cost saving associated with sending less data over switched telephone networks where cost of call is based upon its duration.

- It not only reduces storage requirements

II. Image Compression Methods

The image compression methods are broadly classified into two categories [2] depending whether or not an exact replica of the original image could be reconstructed using the compressed image.

These are:

1. Lossless method
2. Lossy method

2.1 Lossless Compression Method
In lossless compression methods, the original image can be perfectly recovered from the compressed (encoded) image. These are also called noiseless since they do not add noise to the signal (image). It is also known as entropy coding since it uses statistics or decomposition methods to eliminate or minimize redundancy. Lossless compression is used only for a few applications with stringent requirements such as medical imaging. Following methods are included in lossless compression:

1. Run length encoding
2. Huffman encoding
3. LZW coding
4. Area coding

2.2 Lossy Compression Method
Lossy schemes provide much higher compression ratios than lossless schemes. Lossy schemes are widely used since the quality of the reconstructed images is adequate for most applications. Here the decompressed image is not identical to the original image, but reasonably close to it.
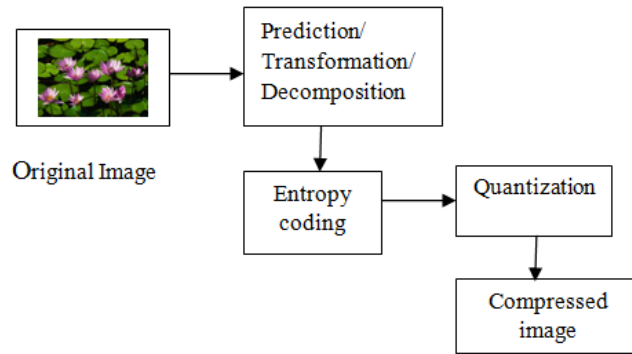
Figure: Outline of lossy image compression

In this 'Prediction – transformation – decomposition' process is completely reversible. The quantization process results in loss of information. The entropy coding after the quantization step, however, is lossless. The decoding is a reverse process. Firstly, entropy decoding is applied to compressed data to get the quantized data. Secondly, dequantization is applied to it & finally the inverse transformation to get the reconstructed image.

Major performance considerations of a lossy compression scheme includes:
1. Compression ratio
2. Signal - to – noise ratio
3. Speed of encoding & decoding.

Lossy compression methods include following schemes:
1. Transformation coding
2. Vector quantization
3. Fractal coding
4. Block Truncation Coding
5. Sub band coding

2.3 Lossless Compression Methods
2.3.1 Run Length Encoding
This is a very simple compression procedure used for sequential data. It is very useful in case of redundant data. This method replaces sequences of identical symbols (pixels), called runs by shorter symbols. [3] The run length code for a gray scale image is represented by a sequence **{Vp ,Rp }** where Vp is the intensity of pixel and Rp refers to the number of consecutive pixels with the intensity **Vp** as shown in the figure.
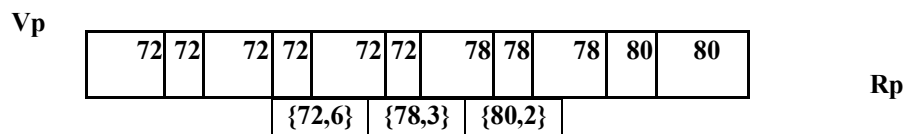

Figure: Run –Length Encoding

2.3.2 Huffman Encoding
This is a common method for coding symbols based on their statistical occurrence of frequencies (probability). [4] The pixels in the image are treated as symbols. The symbols that occur more frequently are assigned a smaller number of bits, while the symbols that occur less frequently are assigned a relatively larger number of bits. Huffman code is a prefix code. This means that the (binary) code of any symbol is not the prefix of the code of any other symbol. Most image coding standards use lossy methods in the earlier stages of compression and use Huffman coding as the last step.

2.3.3 LZW Coding
It is a universal lossless data compression algorithm created by Abraham **L**empel, Jacob **Z**iv, and Terry **W**elch.
LZW   is a dictionary based coding which can be static or dynamic. In static dictionary coding,

dictionary is fixed during the encoding and decoding processes. In dynamic dictionary coding, the dictionary is updated on fly. LZW is widely used in computer industries and it is used as a compress command in UNIX platform.

2.3.4 Area Coding

Area coding is an enhanced version of run length coding, reflecting the two dimensional character of images. This is an important and advance over the other lossless methods. For coding an image it does not make too much sense to interpret it as a sequential stream, as it is in fact an array of sequences, building up a two dimensional object. The algorithms for area coding try to find rectangular regions with the same characteristics. These regions are coded in a descriptive form as an element with two points and a certain structure. This type of coding can be highly effective but it bears the problem of a nonlinear method, which cannot be implemented in hardware. Therefore, the performance in terms of compression time is not competitive, although the compression ratio is.

2.4 LOSSY COMPRESSION METHODS
2.4.1. Transformation Coding

In this coding scheme, transforms such as DFT (Discrete Fourier Transform) and DCT (Discrete Cosine Transform) are used to change the pixels in the original image into frequency domain coefficients (called transform coefficients). [5] These coefficients have several desirable properties. One is the energy compaction property that results in most of the energy of the original data being concentrated in only a few of the significant transform coefficients. This is the basis of achieving the compression. Only those few significant coefficients are selected and the remaining is discarded. The selected coefficients are considered for further quantization and entropy encoding. DCT coding has been the most common approach to transform coding. It is also adopted in the JPEG image compression standard.

2.4.2 Vector Quantization

The basic idea in this method is to develop a dictionary of fixed- size vectors, called code vectors. A vector is usually a block of pixel values. A given image is then partitioned into non-overlapping blocks (vectors) called image vectors then for each value in the dictionary is determined and its index in the dictionary is used as the encoding of the original image vector. Thus, each image is represented by a sequence of indices that can be further entropy coded.

2.4.3 Fractal Coding

The essential idea here is to decompose the image into seg-ments by using standard image processing methods such as color separation, edge detection, and spectrum and texture analysis. [6] Then each segment is looked up in a library of fractals. The library actually contains codes called iterated function system (IFS) codes, which are compact sets of numbers. Using a systematic procedure, a set of codes for a given image are determined, such that when the IFS codes are applied to a suitable set of image blocks yield an image that is a very close approximation of the original. This scheme is highly effective for compressing images that have good regularity and self-similarity.

2.4.4 Block Truncation Coding

In this scheme, the image is divided into non overlapping image of **(M X N)** dimension. **A = [Ø 1 Ø2 … ……ØM]** blocks of pixels. For each block, threshold and reconstruction values are determined. The threshold is usually the mean of the pixel values in the block. Then a bitmap of the block is derived by replacing all pixels whose values are greater than or equal (less than) to the threshold by a 1 or 0. Then for each segment (group of 1s and 0s) in the bitmap, the reconstruction value is determined. This is the average of the values of the corresponding pixels in the original block.

2.4.5 Sub Band Coding

In this method, the image is analyzed to produce the components containing frequencies in well-defined bands, and sub bands. Subsequently, quantization and coding is applied to each of the bands. The advantage of this method is that the quantization and coding well suited for each of the sub bands can be designed separately.

III.   PCA(Principal Component Analysis)

3.1 Introduction

In statistics, PCA is a method for simplifying a dataset by reducing multidimensional datasets to lower dimensions for analysis. PCA is a standard method commonly used for data reduction in statistical pattern recognition and signal processing. PCA is noted to be one of the most valuable results from applied linear algebra. It is used abundantly in all forms of analysis from neuroscience to computer graphics, because it is a simple non - parametric method of extracting relevant information from confusing datasets.

PCA is also called the KARHUNEN-LOEVE Transform (KLT, named after Kari Karhunen & Michel Loeve) or the HOTELLING Transform. Its general objectives are:

1. Data reduction
2. Interpretation

There are basically two approaches for performing PCA. They are Classical Statistical method and Artificial Neural Network method.

3.2 PCA Classical Statistical Method

It involves finding Eigen values and corresponding Eigen vectors of the data set using covariance matrix. [7] The corresponding Eigen values of the matrix gives an indication of amount of information the respective principal components represent. The methodology for calculating principal component is given by the following procedure.

Let X1,X2,……..Xm be the sub images of dimension N. The corresponding algorithm is described as follows:

1. Computation of the global mean (X) from sub images.
2. Subtraction of the mean from each sub image to generate the mean removed image. **Øi = Xi - X**
3. Formation of the matrix using mean removed sub
4. Computation of the sample covariance matrix (C) of dimension **(N x N)**
5. Computation of the Eigen values of the covariance matrix. Computation of Eigen values is performed by Jacobian iteration method.
   C : λ1 > λ2 > ……… >λN
6. Computation of the Eigen vectors for the Eigen values
   C: u1 , u2, …………..,uN
7. Dimensionality reduction step.

Keep only the Eigen vectors corresponding to K largest Eigen values. These Eigen values are called as "principal components". The above said steps are needed to generate the principal components of the image. Corresponding Eigen vectors are uncorrelated and have greater variance. In order to avoid the components that have an undue influence on the analysis, the components are usually coded with mean as zero and variance as one. This standardization of the measurement ensures that they all have equal weight in the analysis.

3.3 PCA Neural Network

Artificial neural network are models that attempt to achieve performance via dense inter connection of simple computational elements. The most important property of a Neural Network is the ability to learn from its environment. PCA is a powerful linear block transform coding in which, an image is subdivided into non-overlapping blocks of N×N pixels which can be considered as N-Dimensional vectors with N = n × n. A linear Transformation, which can be written as an **M ×N – dimensional matrix W with M ≤ N,** is performed on each block with the M rows of W, wi being the basis vectors of the transformation. An Adaptive Principal Component EXtraction (APEX) is used to de-correlate the principal components. The main difference between this APEX architecture and the existing PCA networks lies in the additional lateral connections at the outputs of the network.

3.4 Applications of PCA in Computer Vision Representation

When using these sorts of matrix methods in computer vision, representation of images should be considered. [8] A square, N by N image can be expressed as an N2 – Dimensional vector.

X = (x1 x2 x3 …………………xN2)

Where, the rows of pixels in the image are placed one after the other to form a one dimensional image. Eg. The first N elements x1 – xN will be the first row of the image, the next N elements are the next row, and so

on. The values in the vector are the intensity values of the image, possibly a single grayscale value.

3.4.1 PCA to find patterns

Say we have 25 images with each image is N pixels high by N pixels wide. For each image we can create an image vector as described in the representation section. All these images can be put together in one big image-matrix as given below.

$$\text{Image Matrix} = \begin{bmatrix} \text{ImageVec 1} \\ \text{ImageVec2} \\ \\ \text{ImageVec25} \end{bmatrix}$$

It works much better for recognizing faces, because the PCA analysis has given the original images in terms of the differences and similarities between them. The PCA analysis has identified the statistical patterns in the data.

3.4.2 PCA for Image Compression

If we have 25 images each with N2 vectors and 25 dimensions. Now, PCA can be implemented on this set of data. 25 Eigen vectors will be obtained because each vector is 25 –dimensional. To compress the data, choose the data using only 20 Eigen vectors. This gives a final data set with only 20 dimensions, which has saved one fifth of the space. However, when the original data is reproduced, the images have lost some of the information. This compression is said to be lossy because the decompressed image is not exactly the same as the original.

IV. Conclusion

This paper presents various types of image compression methods. There are basically two types of compression methods. One is Lossless Compression and other is Lossy Compression method. Comparing the performance of compression methods is difficult unless identical data sets and performance measures are used. Some methods perform well for certain classes of data and poorly for other classes of data. The PCA Neural Network approach provides new ways of generating codebook based on statistical feature of PCA. By combining lossy and lossless methods new hybrid semi lossy or semi lossless methods can be created.

References

[1]   Subramanya A,"Image Compression Technique,"Potentials IEEE, Vol.20, Issue 1, pp 19-23, Feb-March 2001, David Jeff Jackson & Sidney Joel Hannah.

[2]   David Jeff Jackson & Sidney Joel Hannah.," Compression   Techniques," System Theory 1993, Proceedings SSST'93 25th o theastem Symposium,pp 513-517, 7-9 March 1993

[3]   Ming Yang Nickolaos Bourbakis ," An Overview of Lossless Digital Image Compression techniques," Circuits & Systems , 2005 48th Midwest Symposium , Vol.2, IEEE ,pp 1099-1102 ,7-10 Aug,2005.

[4]   Ismail Avcibas, Nasir Memon, Bulent Sankur, Khalid Sayood, " AProgressive Lossless / Near Lossless Image Compression Algorithm,"IEEE Signal Processing Letters, vol. 9, No. 10, pp 312-314, October 2002.

[5]   Milos Klima, Karel Fliegel, "Image Compression Techniques in the field of security Technology: Examples and Discussion, "Security Technology, 2004, 38th Annual 2004 Intn. Carnahan Conference, pp 278-284,11-14 Oct., 2004

[6]   Dr. Charles F. Hall, " A Hybrid Image Compression Technique,"Acoustics Speech & Signal Processing, IEEE International Conference on ICASSP' 85, Vol. 10, pp 149- 152, Apr, 1985

[7]   Shi-Fei Ding, Zhong –Zhi Shi,Yong Liang , Feng-Xiang Jin, "   Information Feature Analysis and Improved Algorithm of PCA," Proceedings of the 4th International Conference on Machine Learning and Cybernetics, Guangzhou, pp 1756-1761 , 18-21 August,2005

[8]   Vo Dinh Minh Nhat, Sung Young Lee, "Two-Dimensional Weighted PCA algorithm for Face Recognition," Proceedings 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation, pp 219-223, June 27-30,2005, Espoo, Finland