Communications in Applied Sciences

ISSN 2201-7372

Volume 4, Number 2, 2016, 67-78

Image Compression Based on Data Folding and Principal Component Analysis

Asia Mehdi Nasser, Heba Adnan Raheem

Computer Department, College of Sciences, Kerbala University, Iraq

Corresponding author: Heba Adnan Raheem, Computer Department, College of Sciences, Kerbala University, Iraq

---

Abstract: image compression assumes a fundamental part in image handling field particularly when we need to send the image through a system. While imaging methods produce restrictive measures of information and preparing expansive information is computationally costly, information compression is crucial instrument for capacity and correspondence purposes. Numerous present compression strategies give a high compression rates however with impressive loss of image quality. This paper displays a methodology for image compression in spatial space utilizing an idea of data folding. data folding procedure has been connected on shading images with various size. A row folding is connected on the gray image grid took after by a column folding iteratively till the image size diminishes to predefined esteem as indicated by the levels of folding and unfolding iteration(reconstruction the original image). While Data unfolding process connected in adores mode. Then using *principal component analysis* as a statistical technique concerned with elucidating the covariance structure of a set of variables and uses orthogonal transformation to convert that set of observations of possibly correlated variables into a set of values of linearly uncorrelated and ordered variables called principal components. Method is tested on several standard test images and found that the quality of reconstructed image and compression ratio are ameliorated. The proposed Method is tried on a few standard test images and found that the nature of reproduced image and compression proportion are improved.

***Keywords*—** image compression, data folding, PCA, data unfolding.

© Copyright 2016 the authors.

I. **INTRODUCTION**

Information compression has an essential part in the area of transmission and capacity. It think about as a key in data innovation. The diminishment of redundancies in information representation keeping in mind the end goal to reduction information stockpiling necessity is characterized as information compression which accomplishes a given need Information hypothesis is portrayed as the examination of powerful coding[**1].**Data compression is predicting so as to make a record littler the most continuous bytes and putting away them in less space**[2].**it might be seen as a branch of Information hypothesis in which the fundamental target is to minimize the measure of data to be transmitted. Information compression has a basic part in the scope of transmission and putting away. It expect a key part in information development. The decreasing of redundancies in data representation in order to decay data putting away essential is described as information compression**[3].** It used less utilization of advantages, for instance, memory space or transmission limit. Information compression is named lossless and lossy compression. Lossless compression is used for substance and lossy compression for image**[4].**The rule of image compression calculations is to change paired digits into another that contains the same information yet with less digits, so the record can be as meager as would be reasonable, in this way it imperative in image improvement **[5].** Any lossless coding structure on a very basic level includes three stages. Change, information to-image mapping and lossless image coding. Lossless image doling so as to code delivers a parallel piece stream out twofold code words to the information image **[8].** We suggest an optimal technique to compress the images by using a concept of data folding.

II. RELATED WORKS

In **[7]** the proposed technique introduces a methodology for lossless image compression utilizing idea of data folding for image followed by Huffman encoding. The proposed technique utilizes the property of contiguous neighbor excess for expectation. The closeness between this strategy and our proposition technique in relative execution pressure.

In **[6]** the proposed strategy introduces a methodology for lossless and lossy image compression of image folding followed by Huffman and arithmetic encoding. The outcomes demonstrates that lossless compression is accomplished just at the locale of hobby and it is for the most part suitable for medicinal images.

III. PRINCIPAL COMPONENT ANALYSIS

Principal Component Analysis, or simply PCA, is a useful statistical technique concerned with elucidating the covariance structure of a set of variables and utilizes orthogonal change to convert over that arrangement of perceptions of conceivably associated factors into an arrangement of estimations of directly uncorrelated factors called the main segments [11]. Really, preparing large images is computationally costly by decreasing a large images and as yet saving the basic information in the image that is required for handling, we can accelerate the preparing and utilize less assets [10]. In terms of math, the main step of the standard principal component analysis is to construct a covariance matrix and compute its eigenvalues and eigenvectors. "This process involved representing of N x N image as a one-dimensional vector of N * N elements, by placing the rows of the image one after another.Then we compute the covariance matrix of the entire data set and calculate the eigenvector of this covariance matrix"[9]. The eigenvector with the largest eigenvalue is the direction of greatest variation, the one with the second largest eigenvalue is the orthogonal direction with the next highest variation and so on. The formula to calculate the covariance matrix given in following equation which demonstrates the low computational complexity for large images.

$$C = (AA')_{2x2}$$
$$where\ A = X - M_x$$

X = [x1, x2 . . . xn] represents a set of two dimensional column vectors of chromatic components a* and b* obtained from previous section. We use the median Mx instead of the mean mx to avoid errors caused by outliers and to normalize the data. Now that the data is centered with median 0.0 and standard deviation 1.0, we compute the eigenvalue and eigenvector of the covariance matrix to determine the coefficient matrix which generates the principal components. To solve the eigenvalue λ problem for an n by n matrix, follow steps:

1. Compute the determinant of (C –λ I) with λ^n subtracted along the diagonal, this determinant starts with λ^n or - λ^n, and it is a polynomial in λ of degree n.
2. Find the roots of this polynomial, by solving det (C –λ I) = 0 .The n roots are the n eigenvalues of C. They make C –λ I singular.
3. For each eigenvalue λ, solve (C –λ I) x=0, to find an eigenvector x.

Last, to calculate the principal components themselves, simply multiply the standardized data with the transpose of the eigenvectors of data covariance matrix [11]. "To get the original data back we have to consider all the eigenvectors in our transformation. If we discard some of the less significant eigenvectors in the final transformation, then the retrieved data will lose some information. However, if we choose all the eigenvectors, we can retrieve the original data"[9].

IV. METHODOLOGY AND PROPOSED METHOD

As mentioned in a previously, our work is based on data folding algorithm followed by PCA coding according to following steps we implement data folding :

A. Proposed Solution

1) Input a color image.
2) Convert a color image into a gray scale image.
3) In this step we apply column folding followed by row folding.
4) The PCA coding is fed by the resultant matrix of step 2 to get the compressed image.
5) Masking of PCA coefficients.
6) Save the compressed image in file and compute the compression ratio and PSNR
7) Apply PCA Decoding Algorithm on the compressed image
8) Apply data unfolding algorithm on the resultant image after decompression steps to get the reconstructed image finally.

In the proposed method ,"the compression idea is based on spatial resolution for lossless image compression called *data folding*" [6,7]. The idea is to subtract even pixels from odd pixels and store the difference data in a buffer called *difference matrix*. In this Image compression method first column folding is applied followed by the row folding iteratively till the image size reduces to predefined value. The pixel redundancies are rearranged in a tile format and PCA technique is applied which results the compressed image at the end before transmitting the data. "The goal is to reduce image size iteratively in terms of dimensions - rows or columns by 2".[6].

B. ENCODER

1. Column and Row Folding

In this work, we are used squared image that has a total equaled size. We are applying the folding algorithm on the original color image. "Data folding is an iterative procedure, column folding followed by row folding, that is iterate at each image level. Original image (input image) must be square"[6].

Define a buffer „F (also called picture lattice) whose size is equivalent to that of unique imagee. Unique image is considered as data image for the primary emphasis. At first, the support „F" is empty. In column folding, odd segments of the data image are subtracted from its privilege nearby even segments and put away in first half segments of the unfilled part of the cradle „F". Odd segments are put away in an alternate cradle Odd which is taken as input image to row folding. Following equations depicts column folding technique.**[7].**

F(X+i, Y+j) = gray i, 2j-1) – gray (i, 2j)
Odd (i, j) = gray(i, 2j-1)
i € [1, W] and j € [1, W/2]

Where, gray = input image. W = width of input image, Odd = odd row or column of input image, X = starting x-coordinate of empty portion of matrix F, Y = starting y-coordinate of empty portion F.

"Row folding is like column folding. In row folding, odd lines are subtracted from its contiguous even lines and put away in first half lines of of empty portion of F, Odd lines are put away in an alternate matrix Odd which is taken as an input to next step. Watch that, input image to column folding is always square whereas it is rectangular to row folding".**[7].**

2. System Architecture

The flow charts of compression procedure and data folding as shown in Fig.1, Fig2 and Fig3:

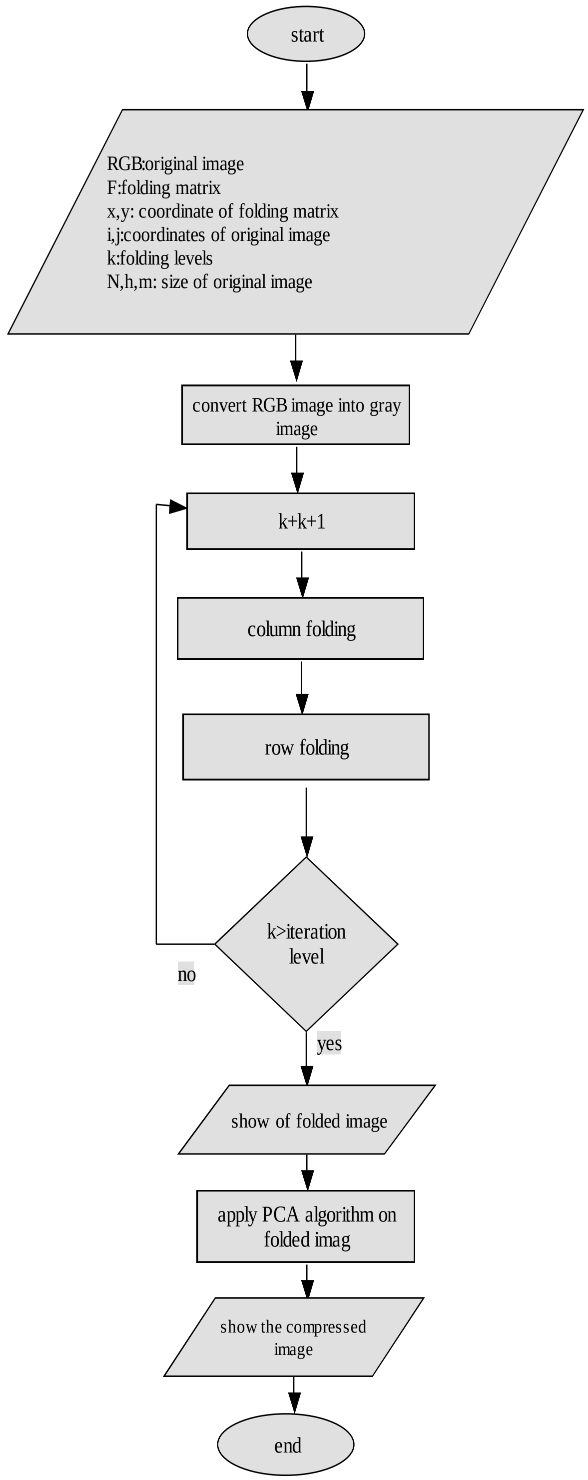

Fig.1 Flowchart of Compression Procedure

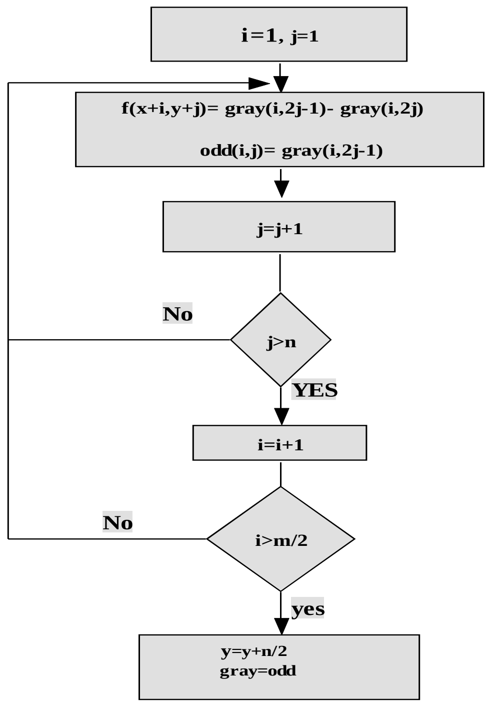

Fig 2.Flowchart of Column Folding

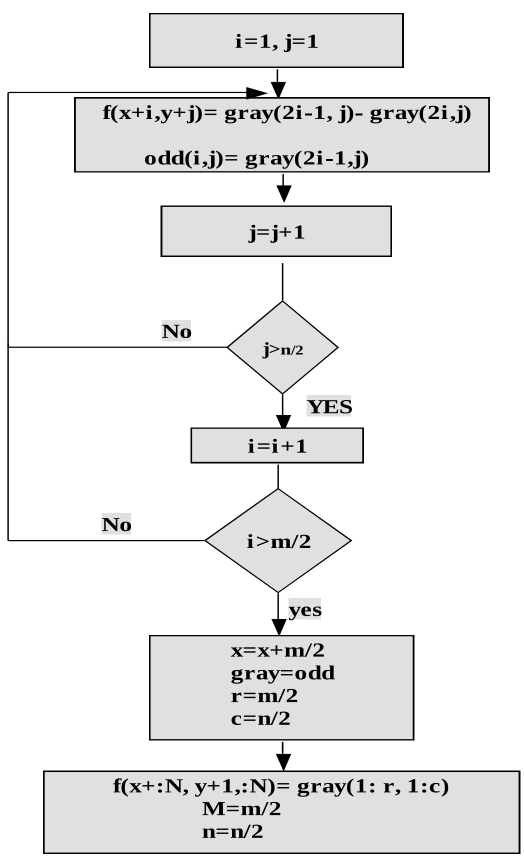

Fig .3 Flowchart of Row Folding

3. Coding

The difference matrix can be encoded by using PCA algorithm (as we mentioned above). The final result obtained after encoding the data of each level would be compressed data for the input image. The compression is measured as the percentage of file size after compression to file size before compression as shown in Following equations:

```
100*fileSize/numel(X);          % Compression Ratio

varargout{2} = (fileSize*8*size(X,3)) /numel(X);  % Bits Per Pixel
```

Where, numel = number of elements in image X.

X= gray scale image.

C. Decoder

The decoder will execute PCA decoding followed by data unfolding. The flow chart of data unfolding as shown in Fig.4

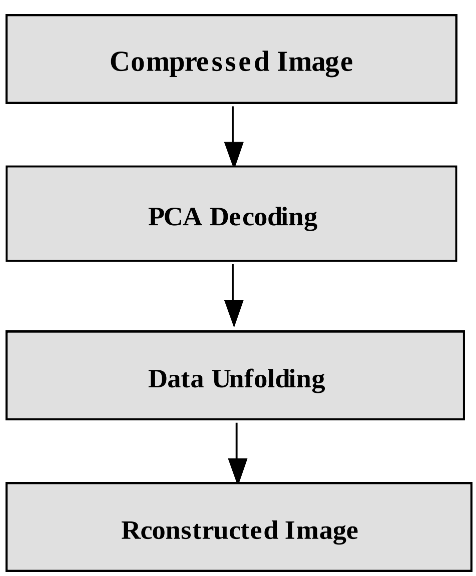

Fig 4: Flow chart of data Unfolding

VI. RESULTS

The proposed compression method is tested on several standard test images such as lena.png, mandril.tif, coloredchips.tif, with diverse sizes such as 256 X 256, 512 X 512, 1024 X 1024, 2048 X 2048. For each image we make it as a square image matrix by changed its spatial coordinates then the number of bits required to store the digitized input image is calculated depended on its spatial resolution, so for a typical 512 X 512 image with 256 levels we need 2,097,152 bits or 262,144 bytes. Then we apply column folding followed by row folding for different levels. Figure (6) depicted the results of applying image folding algorithm for 4 levels.

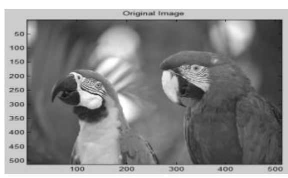

Fig.5 Original Image

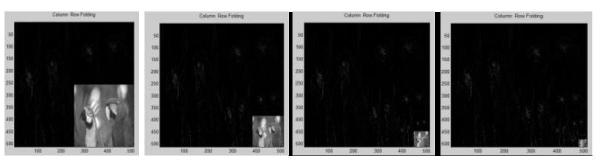

Fig.6 Image Folding Algorithm for 4 Levels

While figure (7) shows the final folded image for four levels and without buffer matrix.

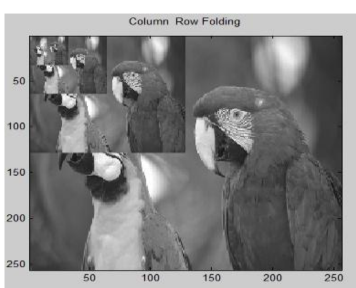

Fig.7 final folded image

In data unfolding of figure (8) showed only row unfolding scheme for four levels without any loss of image quality.

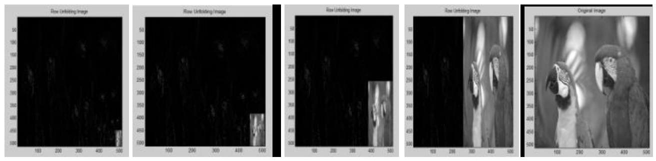

Fig.8 Row Unfolding Image with decompressed image

From the table (1) we notice the variety of file sizes effect the amount of data to be compressed. Also effect on the compression Ratio. and in decompressed part the file size has changed also. Because in compressed part a lot of information added, header data. *we must notice that PC for whole table=50.*

Table(1) show the compression ratio for different size of tested images.

Image name	Image size	level	Bits per pixel (bpp)	Compression ratio
Peppers	1024X1024	2	0.1607	16.0736
Flower	512X512	2	0.2454	24.5361
Purble	256X256	2	0.3149	31.4941
Coloredchips	1024X1024	4	0.3435	34.3506
Purble	256X256	3	0.6406	64.0625
Lena	256X256	2	0.3857	38.5742
Lena	256X256	3	0.7041	70.4102
Coloredchips	512X512	4	0.6826	68.2617
Lena	1024X1024	5	0.6982	69.8242

V. CONCLUSION

The execution of the proposed work in matlab, its worldwide programming language. Where it properties of interfaces backing and be agreeable and well known to the client. The outcomes demonstrated that Data folding is a basic and speedier strategy for compression, and an extremely valuable for security since all the delicate data will be spared in the last pixel of folded image, where the protection of the proposed compression technique depending on folded image, which contain the distinctions of odd positions. It works similarly better for smooth images, in this manner it is more suitable for medicinal images.

REFERENCES

[1] Simrandeep Kaur, Student1; V.Sulochana Verma, Project Consultant, "Design and Implementation of LZW Data Compression Algorithm". International Journal of Information Sciences and Techniques (IJIST) Vol.2, No.4, July 2012.

[2] Arturo San Emeterio Campos, "What's compression, who can be interested and why, and how I'll explain it ". http ://www.arturocampos.com/cp_ch0.html

[3] Abhishek Kumar Singh Hyperlink http://aspalliance.com/author.aspx?uId=65797, an article "Compression and Decompression of Files using Visual Basic 2005", http://aspalliance.com/1287_Compression_and_Decompression_of_Files_using_Visual_Basic_2005. 2.

[4] T. C. Bell, J. G. Cleary, and I. H. Witten. "*Text Compression*". Prentice Hall, Englewood.

[5] Khalid Sayood, "Data Compression", University of Nebraska, Lincoln.

[6] Richa Goyal1, Sourav Garg2 ,"Lossless Image Compression Using Data Folding Followed By Arithmetic Coding", IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 2, Ver. III (Mar – Apr. 2015), PP 67-70 www.iosrjournals.org.

[7] Suresh Yerva, Smita Nair, Krishnan Kutty "Lossless Image Compression based on Data Folding" IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011 MIT, Anna University, Chennai. June 3-5, 2011.

[8] Alan Conrad Bovik, "The Essential guide to Image Processing, vol". 16, pp. 385-419, 2009.

[9] M. Petrou and P. Bosddogianni. *Image Processing: The Fundamentals*. John Wiley and Sons, Inc., UK, 2000. 335 pages, ISBN 0-471-99883-4.

[10] N. Kambhatla and T. K. Leen, "Dimension Reduction by Local Principal Component Analysis," Neural Comput., vol. 9, no. 7, pp. 1493–1516, 1997.

[11] L. I. Smith, "A tutorial on Principal Components Analysis Introduction," *Statistics (Ber).*, vol. 51, p. 52, 2002.