Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.447-454

Using multidimensional scaling technique in image dimension reduction for satellite image

Asmaa Ghalib AL-Rawi[1], Mohammed Abdul Wadood[2]
[1,2] Baghdad University, College of Administration and Economic, Statistic Department

ABSTRACT

A satellite multispectral sensor provides data in the form of several spectral image of particular area of the earth under observation. The amount of data by multispectral exhibit high-inter band correlation with redundancy of the information. In this research, we suggested using multidimensional scaling technique as a statistical technique to reduce image dimensions for reconstruction of a new image form multispectral image. The results proved the efficiency of this technique in providing high quality low dimension images based on the value of Peak Signal to Noise Ratio (PSNR) that was measured to the new image. At the same time, the ability of this technique to reduce dimensions while preserving the basic characteristics of the image has been confirmed.

Keywords: Image processing, Dimension Reduction, Multivariate Statistical Analysis, Multidimensional Scaling

Corresponding Author:

Asmaa Ghalib AL-Rawi
College of Administration and Economic, Statistic Department,
Baghdad University, Iraq
Email: drasmaa.ghalib@coadec.uobaghdad.edu.iq

1. Introduction

Image processing is a technique that deals with mathematical operations on images in order to improve it according to specific criteria. These criteria search every image with extracted useful information by using computer algorithms in a way that gives the best utility for image interpreting. It is the most useful and popular instrument for transmitting data. It enriches thousands of words and briefly convey information about the locations, sizes, and interrelationships between organisms, where About (75%) of the information received by humans is in the form of visual information [3]. It is a tool that describes the visual perception reached by human to get through a form similar to something that is usually a physical object or a person. Despite the importance of the human eye, we need computers that are indispensable in the analysis of large amounts of data and also for tasks that require complex calculations. Because the image is with matrix of pixels, and has multiple dimensions and spectra and the eye is unable to recognize or see these spectra [4]. Also, because of the evolution of devices and applications based on principle of photography, which enabled us to take a large number of images for the same scene. For example, the development of satellites enabled us to take more than a multi-spectral image for a particular area. This large number of taken images to study a specific goal creates difficulty in interpretation and leads to confusion even on human vision when we are trying to analyze for the purpose of extracting information from the image. Therefore, in such cases, the use of multivariate statistical analysis methods for the purpose of reducing large numbers of images to a smaller number is the most appropriate solution to address the problem of multiformity.
Multivariate statistical analysis is the analysis of the data that's in the form of groups of scales related to a number of elements or goals. It has the ability to analyze a large number of associated variables that it adopts

[5]. So, the problem of research is the number of taken images for the studied area and the different spectra of these images and thus the dimensions of each image which will lead to a surplus in the amount of information to be analyzed for the purpose of studying this surplus. In turn, it leads to confusion and difficulty in determining the studied scene, and the importance of this research lies in the fact that it seeks to shed light on a topic that's one of the most important and modern topics for image processing using multivariate statistical methods. We have found a state of compatibility between the science or methods of statistics with computer methods due to importance of statistical methods which enable us to solve various problems with the least effort and time and cost possible.

The research aims to analyze and process digital images by using multidimensional scaling technique as a nonlinear technique to reduce images dimensions. This research was based on a multispectral satellite image of Tigris and Euphrates Rivers and its surrounding areas in southern Iraq equipped with the Iraqi Space Research Center. This region is considered a fertile agricultural area because of containment of channels irrigation, rivers and water bodies scattered at the site in addition to population areas and agricultural land. We get an environmental diversity reflected on the pixels lighting levels in the various color channels of the taken images, which shows how effective of studied methods with such images have different illumination levels.

2. Theoretical side

2.1 Multidimensional scaling

Multidimensional Scaling technique is one of the multivariate statistical analysis methods. It is a statistical technique based on mathematical methodology in its computational steps for the purpose of reducing the databases [5], its seeking to clarify or identifies undiscovered or hidden structure or dimensions for studied problem by studying the correlation between the desired goal variables [6], by studying the similarities (or differences) between a pair of variables and convert it into a distance between points in a reduced dimensions space [6]. These variables, for example, may be the economic indicators of the rise and fall of market prices or the common characteristics of a biomes group. Or, as in this paper for the pixels of the taken digital images, multidimensional scaling technique presents the new reduced space in the form of points closer to each other as the distances between the variables are closer. In this research, the main objective of multidimensional scaling technique is to reduce the surplus of information in the digital image by adopting the distance between the pixels as a criterion for similarity (or difference) within the technique computational methodology, and since the digital image is a three-dimensional image representing a pixel's matrix consisting of association three basic color channels images. We seek to adopt a new mechanism to implement the algorithm of adopted technique to reducing the dimensions that caused a new reduced dimensions image, while preserving the basic features or maintain their color channels of image. We separated the three basic color channels of image as a first stage. As a second stage, we applied the dimension reduction technique on each channel and we get three reduced dimensions images. In the third and final stage, we seek to integrate the resulting reduced dimensions color channels together to form a new reduced three-dimensional image based on mechanism of adopted technique in reduction, has maintained its color features. The importance of this mechanism is to distinguish it from other mechanisms that implementing dimensional reduction algorithms. It studies all the available variables for the purpose of reducing the surplus in order to avoid the reduction of same variables inadvertently. The steps of adopted technique can be explained as follows:

Step 1: Insert the adopted image.

Step 2: Divide the image into its basic color channels (RED (**R**), GREEN (**G**), BULE (**B**)).

Step 3: Transform the image of each color channel into a set of vectors ($\underline{x}_i$) with dimension (N * 1) by transforming each row into a new vector, i.e. the vector ($\underline{x}_1$) consists of (N) pixels of the first row, the second vector ($\underline{x}_2$) takes (N) of pixels of second row, and so on [1].

Step 4: Calculate the arithmetic mean (m_x) for the resulting vectors in the previous step and for each color channel by formula [1]:

$$m_x = E(\underline{x}) = \frac{1}{M}\sum_{i=1}^{N}\underline{x}_i \tag{1}$$

Where ($E(\,)$) is the value of the expectation.
Step 5: Calculate the Euclidean distance matrix (*D*) for the previous vectors and for each color channel [5], by formula:

$$D = \sqrt{\sum_{i=1}^{N}(\underline{x}_{1i} - \underline{x}_{2i})^2} \tag{2}$$

Where ($\underline{x}_1$), ($\underline{x}_2$) are pixel vectors and (i=1, 2, ..., N) represents the number of pixels which consists of each vector.
Step 6: If the distance matrix (D) is not standard, we calculate a matrix (B) according to formula [8]:

$$B = -\frac{1}{2}\,(JD^2J) \;\;\&\;\; J = \left(I - \frac{1}{n}\underline{1}\,\underline{1}^T\right) \tag{3}$$

Where (I) is unity matrix and ($\underline{1}$) is vector have dimension (N * 1) and (n) is the degree of matrix (D), then the dimensions of matrix (J) are (*N* * *N*), and (B) is (N*N).
Step 7: Calculate the eigen values and corresponding eigen vectors for matrix (*B*) [8].
Step 8: Construct conversion matrices (*A*) and (*E*) [8].
Step 9: Extract the new reduced dimension system by making the required conversion according to the following formula [8]:

$$Y = E * A^{\frac{1}{2}} \tag{4}$$

Step 10: If the distance matrix (*D*) is standard, move on to the next step.
Step 11: Calculate the eigen values and eigen vectors of each resulted distance matrix for each color channel.
Step 12: Construct conversion matrix (*A*) according to the results of each color channel [1].
Step 13: Extract the conversion vector ($\underline{y}_i$) according to the following formula [1]:

$$\underline{y}_i = A(\underline{x}_i - m_x) \tag{5}$$

Where (A) is the conversion matrix based on the eigen values and eigen vectors extracted from the matrix (D), ($\underline{x}_i$) are the vectors that formed from the image, (m_x) are the arithmetic means of these vectors.
Step 14: Extract new reduced dimension system based on the vectors extracted in the previous step according to the following equation and for each color channel [1]:

$$\underline{\hat{x}} = A^T\underline{y} + \underline{m}_x \tag{6}$$

Where ($\underline{\hat{x}}_i$) represents the resulting vector set, which represents the reduced dimension system, (A^T) is the inverse of conversion matrix, ($\underline{y}_i$) is the converted vector set in the previous step, (m_x) are the arithmetic means of basic ($\underline{x}_i$) vectors.
Step 15: Combine the new reduced color channels images together.
Step 16: Display the resulting final image.

2.2. Image quality scale

After obtaining the resulting image, in order to know the quality of this image, a statistical scale was adopted to measure the quality of the image after the operations were carried out. This scale is Peak Signal to Noise

Ratio (**PSNR**), which is the difference between the two image pixels [2], and its calculated according to the following formula:

$$PSNR(X,Y) = 10 * log_{10}\frac{(MAX_{pixels}^{2})}{(MSE)} \qquad (7)$$

Where (X) the original image and (Y) is the resulting image after implementation the processing operations, and (MAX) is highest light value for image pixels which is usually (255), and (MSE) is the mean squared error between the adopted and resulting images, where its calculated according to the following formula:

$$MSE = \frac{\left(\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(X_{(i,j)} - Y_{(i,j)}\right)^{2}\right)}{M * N} \qquad (8)$$

Where (M) and (N) are the pixels in the rows and columns. Thus, the higher (**PSNR**) value than zero, indicates an increase in image quality. The values of this criterion are measured by decibel (DB), which it's a logarithmic unit that gives the ratio between two physical quantities, such as the intensity of the sound, the intensity of the light, etc.

3. Application side

According to mechanism of implementation for adopted dimension reduction technique algorithm, we obtained the resulting image (**C**) after reducing the dimensions for studied image (**A**) as shown in Figure 1. Also, in order to clarify how to reduce the dimensions, we extracted the statistical indicators of three basic color channel images which form the new image after the reduction process, and the results are presented in Table 1.

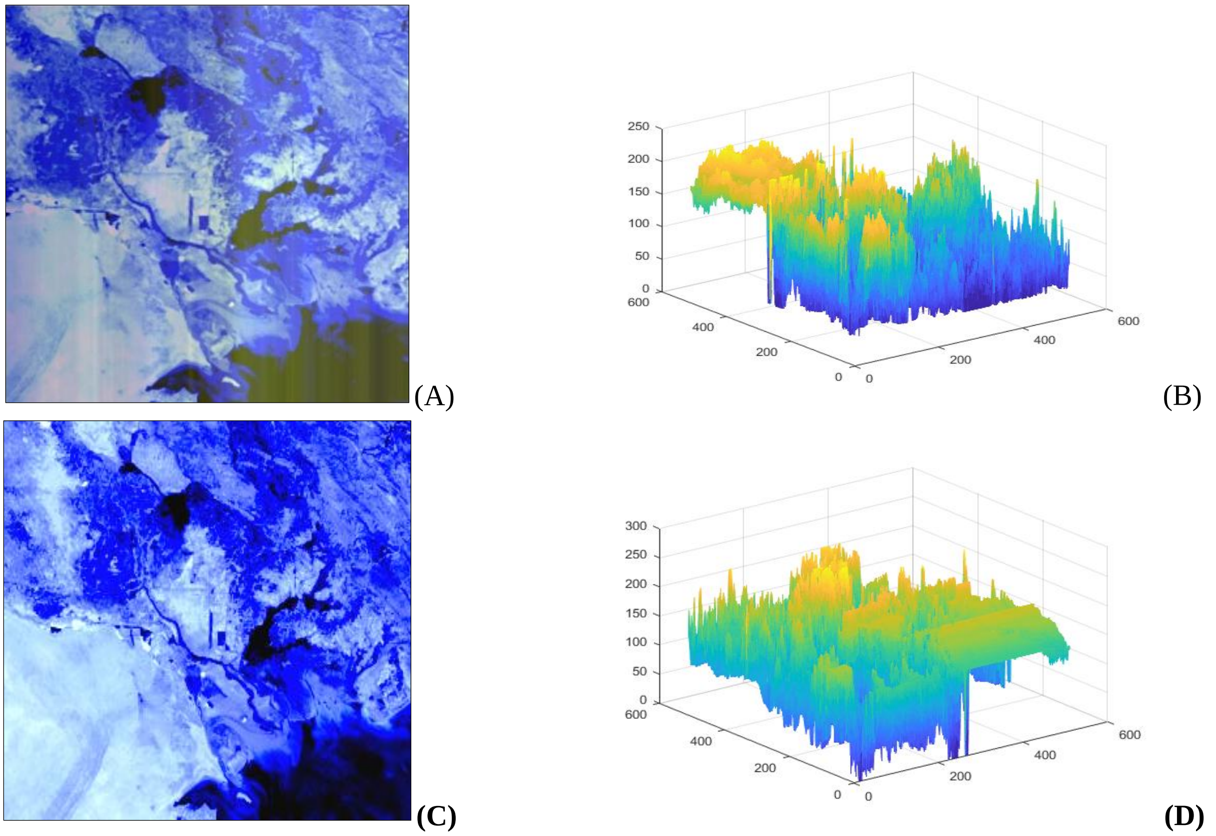

Figure 1. (A) The adopted image before reduction of dimensions; (B) The histogram of the adopted image; (C) the resulting image after dimensional reduction process by multidimensional scaling technique; (D) The histogram of the resulting image

Table 1. The contribution of the highest Eigen values for reduced amount of information for red color channel

λ_i		Total variance for higher Eigen values	Contribution ratio of each Eigen values from total contrast of the image	Percentage Contribution Ratio
1	2.8774	4.5139	0.5014	50.14
2	0.9817		0.1710	17.10
3	0.6548		0.1141	11.41
Total		5.7376	0.7838	78.38

The process of obtaining distance matrix (D) is one of the most important steps for multidimensional scaling technique. It seeks to extract a set of eigen values that reduce most of the visual information and achieve the requirement of reduction and truncate, which represents the proportion of contrast of the main component from the total variance, or which requires at least (75%) of the information to be available in a certain number of the eigen values. The surplus information that is represented by the remaining values can be truncated if we want to reduce the size of the image. The main component that its variance explains the greatest possible from total variation is kept. The components that differ for the total variance are ignored. Some researchers determined the interpretation ratio for the main component from the total variance in the range (70-80%). Others considered that the minimum of the calculated aggregate ratio is equal to (75%) which is calculated according to the formula $\left(\sum_{j=1}^{P}\frac{\lambda_j}{\sum_{j=1}^{p}\lambda_j}\right)$[7], where ($\lambda_j$)is the eigen values. Therefore, the table above is a part of the total table that consists of (512) eigen values equal to number of rows and columns for studied color channel. The table explains the highest eigen values, which meet the statistical requirements for reduction and truncate. So, in red color channel, after the reduction process, the highest three main components were reduced (78.67%) of the total variance, which was calculated from the total variance of three components that equal to (4.5139), divided by total variance that equals to (5.7376). The first eigen value contributed by (50%) of the total variation, followed by second eigen value (17%), then the third (11.14%). It is also noted through the results in the table that these three main components have accumulated amount that equals to (78.38%) of the variance. This refers to the red color channel, while the green channel is based on the following:

Table 2. The contribution of the highest eigen values for reduced amount of information for green color channel

(λ_j)		Total variance for higher eigen values	Contribution ratio of each Eigen values from total contrast of the image	Percentage Contribution Ratio
1	3.7121	5.946	0.4892	48.92
2	1.0984		0.1447	14.47
3	0.7690		0.1013	10.13
4	0.3665		0.0482	4.82
Total		7.588	0.7834	78.34

According to results of this channel, the highest four eigen values had reduced most of the variance that equal to (5.946%), where the first eigen value contributed most of the variance that equals to (48.2%). These four eigen values accumulated an average rate that equal to (78.34%) of total variance for green color channel, while the results of the last color channel (blue channel) have been as following:

Table 3. Shows the contribution of the highest eigen values for reduced amount of information for blue color channel

(λ_j)		Total variance for higher Eigen values	Contribution ratio of each Eigen values from total contrast of the image	Percentage Contribution Ratio
1	3.1415	5.2532	0.4974	49.74
2	1.2141		0.1922	19.22
3	0.8976		0.1421	14.21
Total		6.3146	0.8317	83.17

Where the total variance for this channel is equal to (6.3146) and only three eigen values can be reduced (83%) of the total variation of this channel, which reflects importance of this channel in composition basic image. The total variance of these three values is equal to (5.2532) and first eigen value alone has been reduced with (50%) of the total color channel variation. It is also noted that these three values have collected (83.17%) of the variance and displayed in the last column of Table 3, which reflects importance of this channel in containment most of image information. Therefore, it is concluded that adopted dimensions reduction technique is efficient in increasing the homogeneity of databases in addition to task of reducing the dimensions, which became clear through the results of reducing the dimensions based on eigen values. It is also possible to prove the ability of this technique to give good images that have homogeneity pixels by calculating the difference between arithmetic mean and standard deviation calculated for image and its color channels before and after dimensions reduction. Whenever the difference between these two indicators have increased, the heterogeneity between the pixels has increased [5]. Therefore, when the differences between these two indicators is extracted, we observe increasing pixels homogeneity of image and their color channels after the dimensions reduction, because of approximate these pixels in their intensity and the low difference between the arithmetic mean and standard division that calculated to the image and its color channels before the reduction process. The results are presented in the following table.

Table 4. The statistical indicators for pixels of adopted image and each color channel before and after dimensions reduction process

	Dimension reduction operation							
	Image		Red Channel		Green Channel		Blue Channel	
	Before	After	Before	After	Before	After	Before	After
Means	131.5	129.5	102.9	110.1	129.9	127.1	214.2	193.4
Standard Deviation	62.1	41.6	56.9	41.1	72.2	64.2	55.6	53.8
Differences	69.4	87.9	46	69	57.7	80.9	158.6	139.6

This change in statistical indicators before implementation of multidimensional scaling technique proves that the process of dimensions reducing increases image pixels homogeneity in addition to reducing the dimensions. Thus, it improves the quality of resulting images due to the low proportion of dispersion for image pixels due to reduction of standard deviation value to 41.6 and after that to 62.1. Finally, to measure the quality of resulting image for each color channel, and to see how the resulting image is improved after dimension reducing from the original image, we use the Peak Signal to Noise Ratio (PSNR) shown in paragraph (2.2) and the results are shown in the following table.

Table 5**.** PSNR values for resulting image and color channels before and after dimensions reduction

	Quality indicator before dimensions reduction	Quality indicator after dimensions reduction
Image	10.8402	17.0839
Red Channel	11.3660	17.4126

	Quality indicator before dimensions reduction	Quality indicator after dimensions reduction
Green Channel	9.8266	15.4181
Blue Channel	11.5402	19.2632

According to the results, we note that the quality of the resulting images for color channels and also for image after integration these channels was improved significantly than before the reduction. PSNR for the image after dimension reduction was 17.0839 while, before the reduction, it was 10.8402. This improvement is due to increased quality of color channels images after dimensions reduction than it was. The quality of red, green and blue color channels was 17.41, 15.41 and 19.26 respectively. This is also indicative of the ability of multidimensional scaling technique to produce high-quality images or improve their quality by reducing the surplus in their image information.

4. Conclusions and recommendations

4.1 Conclusions

1. The results showed that dependence on dimension reduction for image color channels gives better results and more obvious than other methods that based on exclusion of effect of color channel. The mechanism adopted in implementation of adopted technique algorithm enabled us to study a large-scale spectral image and produce three-dimensional images unlike other reduction methods.
2. Multidimensional scaling technique has proven its quality in giving clearer images according to the (PSNR) value of the resulting image. It is also a good technique because adoption convergence parameters such as Euclidean distance, which in turn is able to take into, account the disparity in pixels lighting and the extent differences between lighting of these pixels.
3. Multidimensional scaling technique has proved its ability to increase pixels homogeneity of resulting images. It was also characterized by its ability to reduce image sizes in addition to its main task of reducing dimensions, while preserving the most prominent statistical and color characteristics image before reducing the size and dimensions.

4.2. Recommendations

1. Adopt other methods to reduce dimensions such as Canonical Correlation or Kernel Principal Component or Symmetrical Analysis and try to compare the results as linear and nonlinear methods to reduce the dimension.
2. Adopt a different dimension reduction mechanism, such as reducing dimensions of rows and columns of the pixels matrix while retaining the color channels.

5. References

[1] A.M. H. AL-Naser and A. G. Al-Rawi, "Using Some of Multivariate Analysis Methods for the Reduction of Imaging Dimensionality in Biological Sciences Applications" Iraqi Journal of Statistical Science, vol.6, no.10, pp.1-29,2006.

[2] A. Dass, and R. K. Shial and B.S. Gouda, "Improvising MSN and PSNR for Finger-Print Image Noised by GAUSSIAN and SALT & PEPPER" Journal of Multimedia & Its Applications, Vol.4, No.4, pp.59-72, 2012.

[3] K. Minakshi , "Digital Image Processing" Journal of Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, Vol.36, No.1, pp.81-102, 2003.

[4] R. C. Gonzalez, R. E. Woods and S. L. Eddins, "Digital Image Processing Using Matlab" Pearson Prentice Hall, Upper Saddle River, New Jersey, USA,2004.

[5] R. Johnson and D. Wichern, "Applied Multivariate Statistical Analysis" 6th Ed., Pearson Prentice Hall, Upper Saddle River, New Jersey, USA, 2007.

[6] R. Pinkley, M. Gelfand, and L. Duan, "When, Where and How: The Use of Multidimensional Scaling Methods in the Study of Negotiation and Social Conflict" Journal of International Negotiation, Vol.10, No.1, pp.79-96, 2005.

[7] T. Salih, Wasfi., M .Ahmad, Rizgar, "Using Principal Component Analysis to Extract Information and Texture Enhancement in Digital Image Processing" Journal of Kirkuk University for Administration and Econometric Science, Vol.2, No.1, pp.154-182, 2012.

[8] W. Martinez and M. Martinez, "Exploratory Data Analysis with MATLAB" Chapman & Hall, CRC Press, USA,2005.