Methodology of Reducing Image Using Principal Component Analysis With Singular Value Decomposition

Sneha NP[1], MS Sunitha Patel[2], Sneha CR[3]
[1, 2, 3] Assistant Professor Dept of Computer Science and Engineering
[1, 2, 3] ATMECE, Mysuru,Karnataka,India-570028

Abstract- *PCA is a useful statistical technique that has found application in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. Singular value decomposition approach is a principal component analysis method, in which single image , are used to describe the reduction in the image. approach seems to be an adequate method to be used in dimension reduction due to its simplicity, speed and learning capability.*

Keywords- Principal Component Analysis (PCA), Singular Value Decomposition(SVD).

I. INTRODUCTION

PCA is an unsupervised technique used to convert from high dimension to lower dimension. PCA is used to find the principal component in the data. Principal component it tells about the relationship between the variables present in the data.PCA is used to identify the hidden data present in the patterns. It also finds the difference and similarities between the data. In a high dimensional data it is very difficult to find the patterns present in such case PCA is a powerful tool for analysis the pattern. It also identifies the patterns and compresses it without loss of much information. . PCA identifies the pattern present in the dataset and projects it to new data by calculating eigenvectors and eigenvalues. It finds the maximum variance and where the data is most spread. Calculate the covariance matrix from which eigenvectors and eigenvalue are calculated. Covariance matrix finds the maximum variance where the data is most spread and also reduces the redundancy. PCA tells about the eigenvector and eigenvalue, the eigenvector is always associated with the eigenvalue. The eigenvalue tells about the length and the magnitude of the eigenvectors. It sorts the eigenvalue from highest to lowest. If the eigen value and eigenvector have similar magnitude and length then it indicates good indicator.

If the eigen value are greater than the eigenvectors then it contains more information like how the data is spread. If the eigenvectors and eigenvalue is equal to zero then it contains less information. Eigenvectors of the covariance matrix tells about the direction in which the data varies the most. The first eigenvectors is the direction in which the data varies the most. The second eigenvectors should be orthogonal to the first eigenvector. The main aim of the PCA is to find the data and it replaces it has new variable but it contains all the information from the original data but it removes the repeated data and data which is not important .Only the most important data will be listed. Calculate the mean of the variable it also calculate the standard deviation which tells about the data how it is spread. Covariance is used to measure the dimension between the variables. The first principal component has the highest variance. The eigenvectors calculated from the covariance matrix should be always orthogonal. It reduces the noise present in the image based on the variance. The data obtained after the PCA should have low or zero correlation between the variables. Eigenvalue tells the number of variance in which direction it is spread most. The eigenvector with the highest number of eigenvalue is called the principal component.PCA selects the principal component with the highest variance.

II. RELATED WORK

In paper [1] a temporal association rule mining approach named T-patterns, applied on highly challenging floating train data. The aim is to discover temporal associations between pairs of time stamped alarms, called events, which can predict the occurrence of severe failures within a complex bursty environment. The main advantage is Association Rule Mining Produces efficient cluster wise data maintenance and main drawback of this paper is cluster set is too large to handle.

In paper [2] the basic objective of feature subset selection is to reduce the dimensionality of the problem while retaining the most discriminatory information necessary for accurate classification. Thus it is necessary to evaluate feature subsets for their ability to discriminate different classes of pattern. Now the fact that “two best features do not comprise the best feature subset of two features” demands evaluation of

all possible subset of features to find out the best feature subset. If the number of features increases, the number of possible feature subsets grows exponentially leading to a combinatorial optimization problem. The advantage of this paper is optimal subset is created, so less time is required for manipulation and drawbacks are data security is an issue, high possibility for data loss.

In paper [3] a new population-based feature selection method that utilizes dependency between features to guide the search. For the particular problem of feature selection, population based methods aim to produce better "or fitter" future generations that contain more informative subsets of features. It is well-known that feature subset selection is a very challenging optimization problem, especially when dealing with datasets that contain large number of features. The main advantage of this paper is applying subset classification algorithm to simplify data and drawback is, it requires more time to manipulate data.

III.ARCHITECTURE OF PROPOSED SYSTEM

3.1 IMAGE REPRESENTATION

A two dimensional digital image can be represented as function I(x,y) where x and y are discrete coordinate quantities. For simple notational clarity and convenience, integer values are used for these discrete coordinates. The origin (0,0) is located on the top left of an image. The coordinate convention used to represent a digital image in this project is shown in figure 1.1

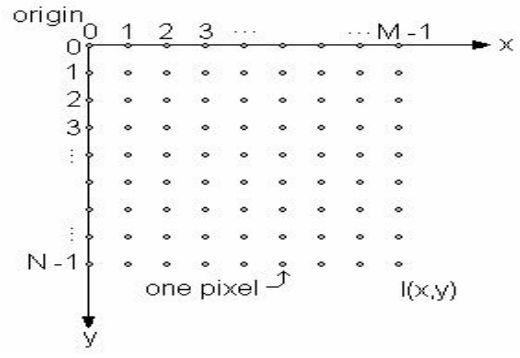

Fig. 1: Standard Coordinates Used to Represent Digital Images.

The representation of images can be written as a N x M matrix on the form

$$I(x,y)=\begin{bmatrix} I(0,0) & I(0,1) & \cdots & I(0,M-1) \\ I(1,0) & I(1,1) & \cdots & I(1,M-1) \\ \vdots & \vdots & & \vdots \\ I(N-1,0) & I(N-1,1) & \cdots & I(N-1,M-1) \end{bmatrix} = \begin{bmatrix} I_{0,0} & I_{0,1} & \cdots & I_{0,M-1} \\ I_{1,0} & I_{1,1} & & I_{1,M-1} \\ \vdots & & & \vdots \\ I_{N-1,0} & I_{N-1,1} & \cdots & I_{N-1,M-1} \end{bmatrix}.$$

Each element of the matrix as shown in Eq. (1.1) is called an image element, picture element, pixel or pel. The terms image and pixel will be used throughout the rest of this project to denote a digital image and its elements. The pixels are registered by a digital sensor and may encode color or intensity. In this project we use color images, but they can be converted to a grayscale intensity image by taking the average of the red, green and blue color components.

Apply PCA to reduce the dimension of the data set

1. Obtain the input matrix table.
2. Subtract the mean from the dataset in all the n-dimensions.
3. Calculate the covariance matrix of this mean-subtracted dataset.
4. Calculate the SVD from the covariance matrix
5. Deriving the new data set.

PCA is to take our data and rewrite it in terms of new variables so that our "new data" has all the information from the original data but the redundancy has been removed and it has been organized such that the most important variables are listed first.Principal components are obtained by projecting multivariate data vectors.

The covariance matrix contains scaled sum of squares and cross products of a multivariate data set. Plot the eigenvalues, if the points on the graph tend to level out then these eigenvalues are usually close to zero so they can be ignored.

3.2 Singular value decomposition and principal component analysis

Singular value decomposition(svd) and principal component analysis(PCA) are two eigenvalues methods used to reduce a high dimensional dataset into fewer dimension while retaining important information.PCA uses the svd in its calculation. SVD is a general matrix decomposition method that can be used on any m*n matrix.The eigenvector with the highest eigenvalue is the first principal component of a data set.

IV. EXPERIMENTAL RESULTS

Figure 3 depicts the input image for PCA with SVD

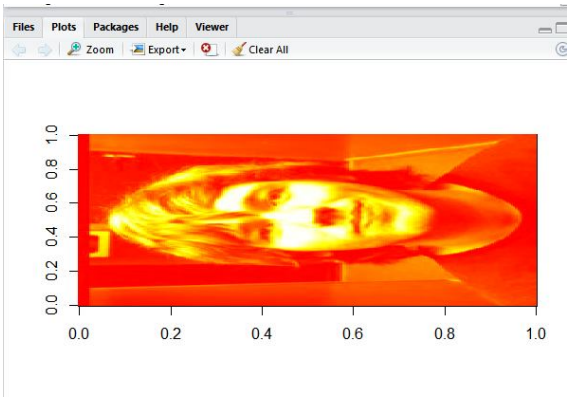

Figure 3: Input Image for PCA with SVD

Figure 4 depicts after applying PCA with SVD

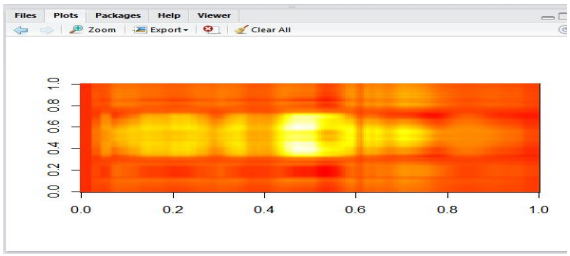

Figure 4:After applying PCA with SVD

V. CONCLUSION

PCA mechanism to eliminate the noise and redundancy in the image.PCA mechanism to transform high dimensional data into lower dimensional data so result improves systematically and significantly.

REFERENCES

[1] Qinbao Song, Jingjie Ni and Guangtao Wang"*A Fast Clustering-Based Feature Subset Selection Algorithm for High Dimensional Data*", IEEE transactions on knowledge and data engineering vol: 25 no: 1 year 2013.

[2] Aastha Joshi., *"A Comparative Analysis of Various Clustering Techniques used for Very Large Datasets"*,International Journal of Advanced Research in Computer Science and Software EngineeringVolume 3, Issue 3, March 2013.

[3] Sammouri W., *"Temporal association rule mining for the preventive diagnosis of onboard subsystems within floating train data framework"*, International journal of engineering research pp 1351-1356, June 2012.

[4] R.Indhumathi.,*"Reducing and Clustering high Dimensional Data through Principal Component Analysis"*, International Journal of Computer Applications (0975 – 8887), Volume 11– No.8, December 2010.

[5] Chakraborty.B.., *"Bio-inspired algorithms for optimal feature subset selection"* In Proceedings of the Fifth IEEE international Conference on Computers and Devices for Communication , pp 1-7, Dec 2012.

[6] M.Vijayalakshmi.,"*A Survey of Different Issue of Different clustering Algorithms Used in Large Data sets*",International Journal of Advanced Research in Computer Science and Software EngineeringVolume 2, Issue 3, March 2012.

[7] Ahmed Al-Ani, Rami N. Khushaba., *"A Population Based Feature Subset Selection Algorithm Guided by Fuzzy Feature Dependency"*,International Journal of Advanced Research Volume 322, pp 430-438, Dec 2012.