Principal Component Analysis - A Tutorial

Alaa Tharwat

Electrical Dept. - Suez Canal University- Egypt

April 2, 2016

- Introduction.
- Principal Component Analysis (PCA).
- Numerical Examples.
- Experiments.
- Conclusions and Future Works.

The main objective is to:

- Explain the Explain how to use PCA in different fields.
- Introduce numerical examples to explain how to apply PCA in different applications.

- The goal of the PCA technique is to find a lower dimensional space or PCA space (W) that is used to transform the data ($X = \{x_1, x_2, \ldots, x_N\}$) from a higher dimensional space ($\mathcal{R}^M$) to a lower dimensional space ($\mathcal{R}^k$), where N represents the total number of samples or observations.

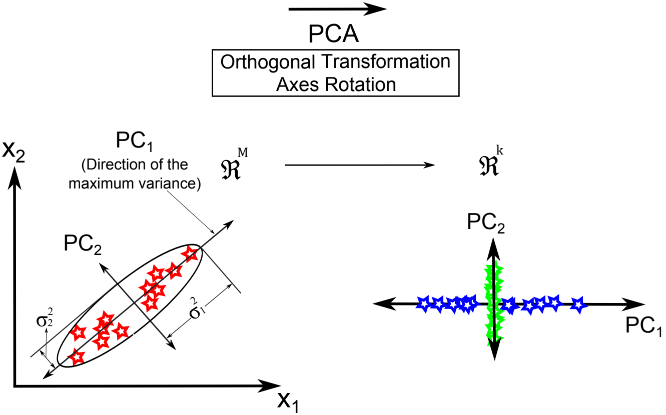


Figure: Example of the two-dimensional data (x_1, x_2).

- The PCA space consists of k principal components. The principal components are orthonormal, uncorrelated, and it represents the direction of the maximum variance.
- The first principal component ($(PC_1$ or $v_1) \in \mathcal{R}^{M\times 1}$) of the PCA space represents the direction of the maximum variance of the data, the second principal component has the second largest variance, and so on.

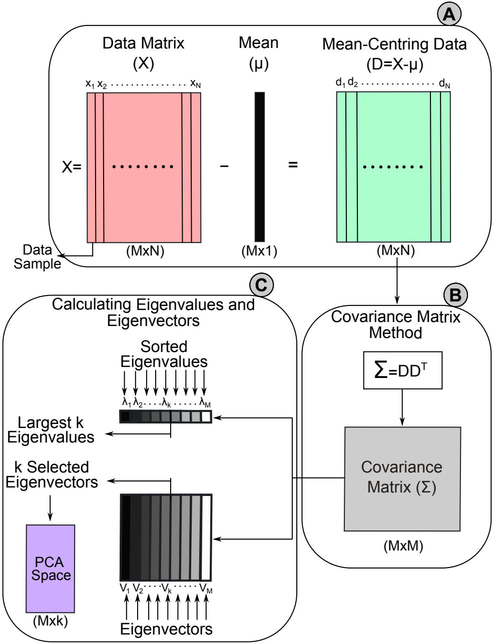
- In this method, there are two main steps to calculate the PCs of the PCA space. First, the covariance matrix of the data matrix (X) is calculated. Second, the eigenvalues and eigenvectors of the covariance matrix are calculated.
- The covariance matrix is used when the number of variables more than one.
- Covariance matrix is a symmetric matrix (i.e. $X = X^T$) and always positive semi-definite matrix.

$$\begin{pmatrix} Var(x_1, x_1) & Cov(x_1, x_2) & \dots & Cov(x_1, x_M) \\ Cov(x_2, x_1) & Var(x_2, x_2) & \dots & Cov(x_2, x_M) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(x_M, x_1) & Cov(x_M, x_2) & & Var(x_M, x_M) \end{pmatrix} \tag{1}$$

- The covariance matrix is solved by calculating the eigenvalues (λ) and eigenvectors (V) as follows:

$$V\Sigma = \lambda V \tag{2}$$

  where V and λ represent the eigenvectors and eigenvalues of the covariance matrix, respectively.
- The eigenvalues are scalar values, while the eigenvectors are non-zero vectors, which represent the principal components, i.e. each eigenvector represents one principal component.
- The eigenvectors represent the directions of the PCA space, and the corresponding eigenvalues represent the scaling factor, length, magnitude, or the robustness of the eigenvectors.
- The eigenvector with the highest eigenvalue represents the first principal component and it has the maximum variance.


A
Data Matrix
(X)
Mean
(μ)
Mean-Centring Data
(D=X-μ)
x1 x2 . . . . xN
d1 d2 . . . . dN
X=
−
=
(MxN)
(Mx1)
(MxN)
Data
Sample
B
Covariance Matrix
Method
Σ=DD^T
Covariance
Matrix (Σ)
(MxM)
C
Calculating Eigenvalues and
Eigenvectors
Sorted
Eigenvalues
λ1 λ2 . . . λk . . . . . λM
Largest k
Eigenvalues
k Selected
Eigenvectors
PCA
Space
(Mxk)
V1 V2 . . . Vk . . . . . VM
Eigenvectors

1: Given a data matrix ($X = [x_1, x_2, \dots, x_N]$), where N represents the total number of samples and x_i represents the i^{th} sample.

2: Compute the mean of all samples as follows:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{3}$$

3: Subtract the mean from all samples as follows:

$$D = \{d_1, d_2, \dots, d_N\} = \sum_{i=1}^{N} x_i - \mu \tag{4}$$

4: Compute the covariance matrix as follows:

$$\Sigma = \frac{1}{N-1} D \times D^T \tag{5}$$

5: Compute the eigenvectors V and eigenvalues λ of the covariance matrix (Σ).

6: Sort eigenvectors according to their corresponding eigenvalues.

7: Select the eigenvectors that have the largest eigenvalues $W = \{v_1, \dots, v_k\}$.

- Singular value decomposition is one of the most important linear algebra principles.
- The aim of the SVD method is to diagonalize the data matrix ($X \in \mathcal{R}^{p \times q}$) into three matrices as in Equation (6).

$$X = LSR^T = \begin{bmatrix} l_1 & \cdots & l_M & \end{bmatrix} \begin{bmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_N \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -r_1^T- \\ -r_2^T- \\ \vdots \\ -r_N^T- \end{bmatrix} \quad (6)$$

where $L(p \times p)$ are called left singular vectors, $S(p \times q)$ is a diagonal matrix represents the singular values that are sorted from high-to-low, i.e. the highest singular value in the upper-left index of S, thus, $s_1 \geq s_2 \geq \cdots \geq s_q \geq 0$, and $R(q \times q)$ represents the right singular vectors. The left and right singular matrices, i.e. L and R, are orthonormal bases.

- The columns of the right singular vectors (R) represent the eigenvectors of X^TX or the principal components of the PCA space, and $s_i^2,\ \forall\ i = 1, 2, \ldots, q$ represent their corresponding eigenvalues

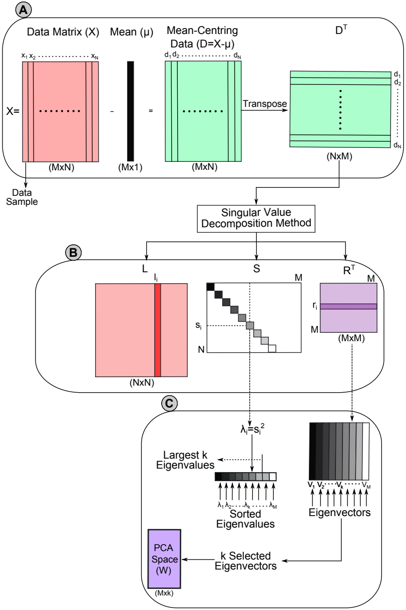


Figure: Visualized steps to calculate the PCA space using SVD method.

1: Given a data matrix ($X = [x_1, x_2, \ldots, x_N]$), where N represents the total number of samples and $x_i(M \times 1)$ represents the i^{th} sample.
2: Compute the mean of all samples as in Equation (3).
3: Subtract the mean from all samples as in Equation (4).
4: Construct a matrix $Z = \frac{1}{\sqrt{N-1}} D^T$, $Z(N \times M)$.
5: Calculate SVD for Z matrix as in Equation (6).
6: The diagonal elements of S represent the square root of the sorted eigenvalues, $\lambda = diag(S^2)$, while the PCs are represented by the columns of R.
7: Select the eigenvectors that have the largest eigenvalues $W = \{R_1, R_2, \ldots, R_k\}$ to construct the PCA space.
8: All samples are projected on the lower dimensional space of PCA (W) as follows, $Y = W^T D$.

- The covariance matrix is the product of DD^T, where $D = \{d_i\}_{i=1}^{N}, d_i = x_i - \mu$. Using Equation (6) that is used to calculate SVD, the covariance matrix can be calculated as follows:

$$DD^T = (LSR^T)^T(LSR^T) = RS^T L^T LSR^T \tag{7}$$

where $L^T L = I$

$$DD^T = RS^2R^T = (SVD(D^T))^2 \tag{8}$$

where S^2 represents the eigenvalues of $D^T D$ or DD^T and the columns of the right singular vector (R) represent the eigenvectors of DD^T.

- To conclude, the square root of the eigenvalues that are calculated using the covariance matrix method are equal to the singular values of SVD method. Moreover, the eigenvectors of Σ are equal to the columns of R. Thus, the eigenvalues and eigenvectors that are calculated using the two methods are equal.

- To construct the lower dimensional space of PCA (W), a linear combination of k selected PCs that have the most k eigenvalues are used to preserve the maximum amount of variance, i.e. preserve the original data, while the other eigenvectors or PCs are neglected.
- The dimension of the original data is reduced by projecting it after subtracting the mean onto the PCA space as in Equation (9).

$$Y = W^T D = \sum_{i=1}^{N} W^T (x_i - \mu) \tag{9}$$

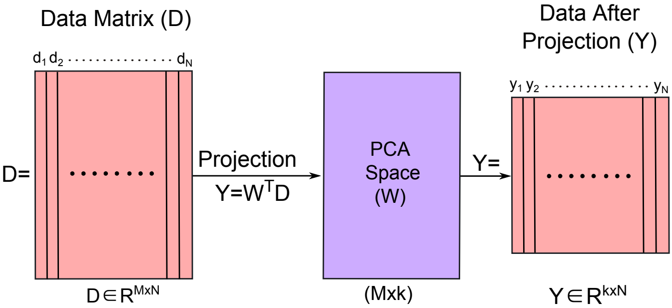


Figure: Data projection in PCA as in Equation (9).

- The original data can be reconstructed again as in Equation (10).

$$\hat{X} = WY + \mu = \sum_{i=1}^{N} W y_i + \mu \tag{10}$$

where $\hat{X}$ represents the reconstructed data.

- The deviation between the original data and the reconstructed data are called the reconstruction error or residuals as denoted in Equation (11).

$$Error = X - \hat{X} = \sum_{i=1}^{N} (x_i - \hat{x_i})^2 \tag{11}$$

- The robustness of the PCA is controlled by the number of selected eigenvectors (k) and it is measured by the sum of the selected eigenvalues, which is called total variance as in Equation (12).

$$\text{Robustness of the PCA space} = \frac{\text{Total Variance of } W}{\text{Total Variance}} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{M} \lambda_i} \tag{12}$$

- Given a data matrix $X = \{x_1, x_2, \ldots, x_8\}$, where x_i represents the i^{th} sample as denoted in Equation (13).

$$X = \begin{bmatrix} 1.00 & 1.00 & 2.00 & 0.00 & 5.00 & 4.00 & 5.00 & 3.00 \\ 3.00 & 2.00 & 3.00 & 3.00 & 4.00 & 5.00 & 5.00 & 4.00 \end{bmatrix} \tag{13}$$

- Each sample of the matrix was represented by one column that consists of two features ($x_i \in \mathcal{R}^2$) to visualize it. The total mean (μ) was then calculated as in Equation (3) and its value was $\mu = \begin{bmatrix} 2.63 \\ 3.63 \end{bmatrix}$.
- The data were then subtracted from the mean as in Equation (4) and the values of D will be as follows:

$$D = \begin{bmatrix} -1.63 & -1.63 & -0.63 & -2.63 & 2.38 & 1.38 & 2.38 & 0.38 \\ -0.63 & -1.63 & -0.63 & -0.63 & 0.38 & 1.38 & 1.38 & 0.38 \end{bmatrix} \tag{14}$$

- The covariance matrix (Σ) were then calculated as in Equation (5). The eigenvalues (λ) and eigenvectors (V) of the covariance matrix were then calculated. The values of the Σ, λ, and V are shown below.

$$\Sigma = \begin{bmatrix} 3.70 & 1.70 \\ 1.70 & 1.13 \end{bmatrix}, \lambda = \begin{bmatrix} 0.28 & 0.00 \\ 0.00 & 4.54 \end{bmatrix}, \text{ and } V = \begin{bmatrix} 0.45 & -0.90 \\ -0.90 & -0.45 \end{bmatrix} \tag{15}$$

- **Calculating the Projected Data**: The mean-centring data (i.e. data – total mean) were then projected on each eigenvector as follows, ($Y_{v1} = v_1^T D$ and $Y_{v2} = v_2^T D$), where Y_{v1} and Y_{v2} represent the projection of the D on the first and second eigenvectors, i.e. v_1 and v_2, respectively. The values of Y_{v1} and Y_{v1} are shown below.

$$\begin{aligned} Y_{v1} &= \begin{bmatrix} -0.16 & 0.73 & 0.28 & -0.61 & 0.72 & -0.62 & -0.18 & -0.17 \end{bmatrix} \\ Y_{v2} &= \begin{bmatrix} 1.73 & 2.18 & 0.84 & 2.63 & -2.29 & -1.84 & -2.74 & -0.50 \end{bmatrix} \end{aligned} \tag{16}$$

- **Calculating the Projected Data**: The reconstruction error between the original data and the reconstructed data using all eigenvectors or all principal components, i.e. there is no information lost, approximately tend to zero.

$$\hat{X}_1 = v_1 Y_{v1} + \mu = \begin{bmatrix} 2.55 & 2.95 & 2.75 & 2.35 & 2.95 & 2.35 & 2.55 & 2.55 \\ 3.77 & 2.97 & 3.37 & 4.17 & 2.98 & 4.18 & 3.78 & 3.78 \end{bmatrix}$$
$$\hat{X}_2 = v_2 Y_{v2} + \mu = \begin{bmatrix} 1.07 & 0.67 & 1.88 & 0.27 & 4.68 & 4.28 & 5.08 & 3.08 \\ 2.85 & 2.66 & 3.25 & 2.46 & 4.65 & 4.45 & 4.84 & 3.85 \end{bmatrix} \tag{17}$$

- The error between the original data and the reconstructed data that were projected on the first and second eigenvectors are denoted by E_{v1} and E_{v2}, respectively. The values of E_{v1} and E_{v2} are as follows:

$$
\begin{aligned}
E_{v1} &= X - \hat{X}_1 \\
&= \begin{bmatrix} -1.55 & -1.95 & -0.75 & -2.35 & 2.05 & 1.65 & 2.45 & 0.45 \\ -0.77 & -0.97 & -0.37 & -1.17 & 1.02 & 0.82 & 1.22 & 0.22 \end{bmatrix} \\
E_{v2} &= X - \hat{X}_2 \\
&= \begin{bmatrix} -0.07 & 0.33 & 0.12 & -0.27 & 0.32 & -0.28 & -0.08 & -0.08 \\ 0.15 & -0.66 & -0.25 & 0.54 & -0.65 & 0.55 & 0.16 & 0.15 \end{bmatrix}
\end{aligned}
\tag{18}
$$

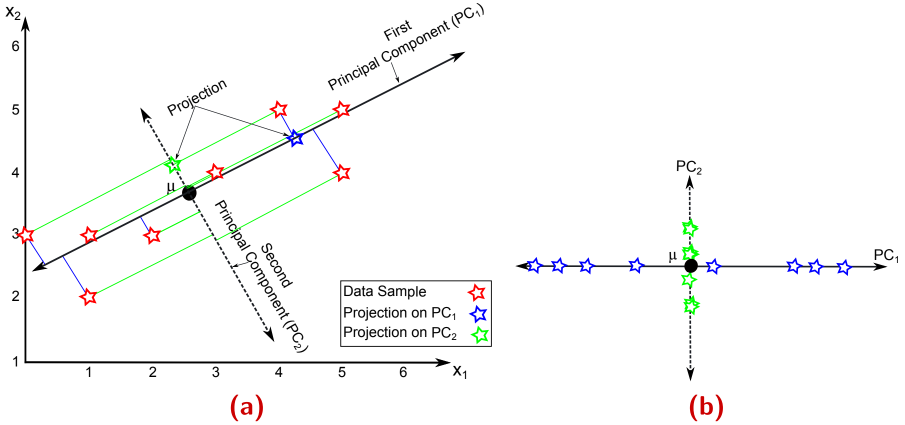


Figure: A visualized example of the PCA technique, (a) the dotted line represents the first eigenvector (v_1), while the solid line represents the second eigenvector (v_2) and the blue and green lines represent the reconstruction error using PC_1 and PC_2, respectively; (b) projection of the data on the principal components, the blue and green stars represent the projection onto the first and second principal components, respectively.

- The first three steps in SVD method and covariance matrix methods are common. In the fourth step in SVD, the original data were transposed as follows, $Z = \frac{1}{N-1}D^T$. The values of Z are as follows:

$$Z = \begin{bmatrix} -0.61 & -0.24 \\ -0.61 & -0.61 \\ -0.24 & -0.24 \\ -0.99 & -0.24 \\ 0.90 & 0.14 \\ 0.52 & 0.52 \\ 0.90 & 0.52 \\ 0.14 & 0.14 \end{bmatrix} \tag{19}$$

- SVD was then used to calculate L, S, and R as in Equation (6) and their values are as follows:

$$
L = \begin{bmatrix}
-0.31 & 0,12 & -0,07 & -0,60 & 0,58 & 0,15 & 0,41 & 0,04 \\
-0.39 & -0,52 & -0,24 & 0,20 & -0,29 & 0,53 & 0,31 & 0,14 \\
-0.15 & -0,20 & 0,96 & -0,01 & -0,01 & 0,08 & 0,07 & 0,02 \\
-0.47 & 0,43 & 0,02 & 0,69 & 0,32 & -0,05 & 0.12 & -0,01 \\
0.41 & -0,51 & -0,04 & 0,31 & 0,68 & 0,08 & -0,09 & 0,02 \\
0.33 & 0.44 & 0,08 & 0,02 & 0,02 & 0,82 & -0,15 & -0,05 \\
0,49 & 0,12 & 0,05 & 0,17 & -0,15 & -0,12 & 0,83 & -0,03 \\
0,09 & 0,12 & 0,02 & 0,00 & 0,01 & -0,05 & -0,04 & 0,99
\end{bmatrix},
$$

$$
S = \begin{bmatrix}
2.13 & 0 \\
0 & 0.53 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0 \\
0 & 0
\end{bmatrix},
$$

$$
R = \begin{bmatrix}
0.90 & -0.45 \\
0.45 & 0.90
\end{bmatrix} \tag{20}
$$

- In this example, each sample was represented by four variables. In this experiment, the second variable was constant for all observations as shown below.

$$X = \begin{bmatrix} 1.00 & 1.00 & 2.00 & 0.00 & 7.00 & 6.00 & 7.00 & 8.00 \\ 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 & 2.00 \\ 5.00 & 6.00 & 5.00 & 9.00 & 1.00 & 2.00 & 1.00 & 4.00 \\ 3.00 & 2.00 & 3.00 & 3.00 & 4.00 & 5.00 & 5.00 & 4.00 \end{bmatrix} \quad (21)$$

- The covariance matrix of the given data was calculated and its values are shown below.

$$\Sigma = \begin{bmatrix} 10.86 & 0 & -7.57 & 2.86 \\ 0 & 0 & 0 & 0 \\ -7.57 & 0 & 7.55 & -2.23 \\ 2.86 & 0 & -2.23 & 1.13 \end{bmatrix} \quad (22)$$

- The eigenvalues and eigenvectors of are

$$\lambda = \begin{bmatrix} 17.75 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.46 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad V = \begin{bmatrix} 0.76 & 0.62 & -0.20 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ -0.61 & 0.79 & 0.10 & 0.00 \\ 0.21 & 0.05 & 0.98 & 0.00 \end{bmatrix} \tag{23}$$

1. The first eigenvector represents the first principal component because it was equal to $\frac{17.75}{17.75+1.46+0.33+0} \approx 90.84\%$ of the total variance.
2. The first three eigenvectors represent 100% of the total variance of the total data and the fourth eigenvector was redundant.
3. The second variable, i.e. second row, will be neglected completely when the data were projected on any of the best three eigenvectors.
4. The projected data on the fourth eigenvector preserved only the second variable and all the other original data were lost, and the reconstruction error was ≈ 136.75, while the reconstruction error was ≈ 12.53 , 126.54 , 134.43 when the data were projected on the first three eigenvectors, respectively,

- The projected data are

$$
\begin{aligned}
Y_{v_1} &= \begin{bmatrix}-2.95 & -3.78 & -2.19 & -6.16 & 4.28 & 3.12 & 4.49 & 3.20\end{bmatrix}\\
Y_{v_2} &= \begin{bmatrix}-1.20 & -0.46 & -0.58 & 1.32 & -0.58 & -0.39 & -0.54 & 2.39\end{bmatrix}\\
Y_{v_3} &= \begin{bmatrix}0.06 & -0.82 & -0.14 & 0.64 & -0.52 & 0.75 & 0.45 & -0.43\end{bmatrix}\\
Y_{v_4} &= \begin{bmatrix}0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00\end{bmatrix}
\end{aligned}
\tag{24}
$$

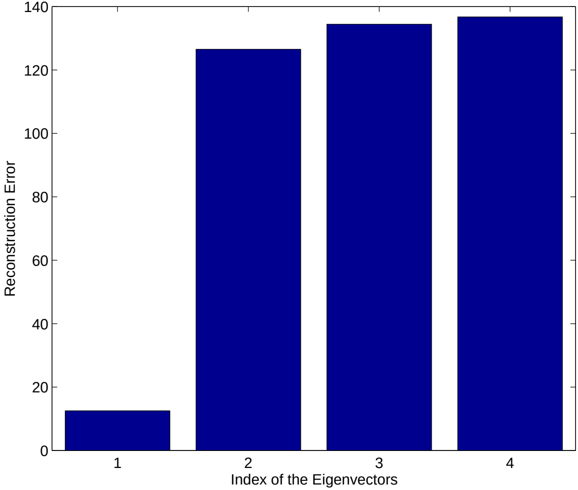
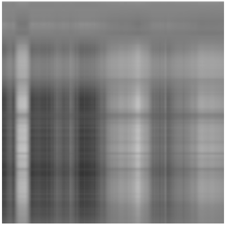


Figure: Reconstruction error of the four eigenvectors in the multi-class example.

Table: A comparison between ORL, Ear, and Yale datasets in terms of accuracy (%), CPU time (sec), and cumulative variance (%) using different number of eigenvectors (biometric experiment).

Number of Eigenvectors	ORL Dataset			Ear Dataset			Yale Dataset		
	Acc. (%)	CPU Time (sec)	Cum. Var. (%)	Acc. (%)	CPU Time (sec)	Cum. Var. (%)	Acc. (%)	CPU Time (sec)	Cum. Var. (%)
1	13.33	0.074	18.88	15.69	0.027	29.06	26.67	0.045	33.93
5	80.83	0.097	50.17	80.39	0.026	66.10	76.00	0.043	72.24
10	94.17	0.115	62.79	90.20	0.024	83.90	81.33	0.042	85.13
15	95.00	0.148	69.16	94.12	0.028	91.89	81.33	0.039	90.18
20	95.83	0.165	73.55	94.12	0.033	91.89	84.00	0.042	93.36
30	95.83	0.231	79.15	94.12	0.033	98.55	85.33	0.061	96.60
40	95.83	0.288	82.99	94.12	0.046	99.60	85.33	0.064	98.22
50	95.83	0.345	85.75	94.12	0.047	100.00	85.33	0.065	99.12
100	95.83	0.814	93.08	94.12	0.061	100.00	85.33	0.091	100.00

Acc. accuracy; Cum. Cumulative; Var. variance.



(a) One principal component, CR=512:1, and MSE=30.7

(b) 10% of the principal components, CR=10:1, and MSE=5.31

(c) 50% of the principal components, CR=2:1, and MSE=0.909

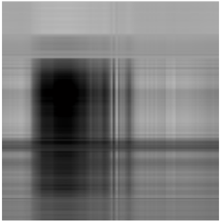


Table: Compression ratio and mean square error of the compressed images using different percentages of the eigenvectors (image compression experiment).

Percentage of the used Eigenvectors	Lena Image			Cameraman Image		
	MSE	CR	Cumulative Variance (%)	MSE	CR	Cumulative Variance (%)
10	5.3100	512:51.2	97.35	8.1057	256:25.6	94.56
20	2.9700	512:102.4	99.25	4.9550	256:51.2	98.14
30	1.8900	512:153.6	99.72	3.3324	256:76.8	99.24
40	1.3000	512:204.8	99.87	2.0781	256:102.4	99.73
50	0.9090	512:256	99.94	1.1926	256:128	99.91
60	0.6020	512:307.2	99.97	0.5588	256:153.6	99.98
70	0.3720	512:358.4	99.99	0.1814	256:179.2	100.00
80	0.1935	512:409.6	100.00	0.0445	256:204.8	100.00
90	0.0636	512:460.8	100.00	0.0096	256:230.4	100.00
100 (All)	0.0000	512:512=1	100.00	0.0000	1	100.00

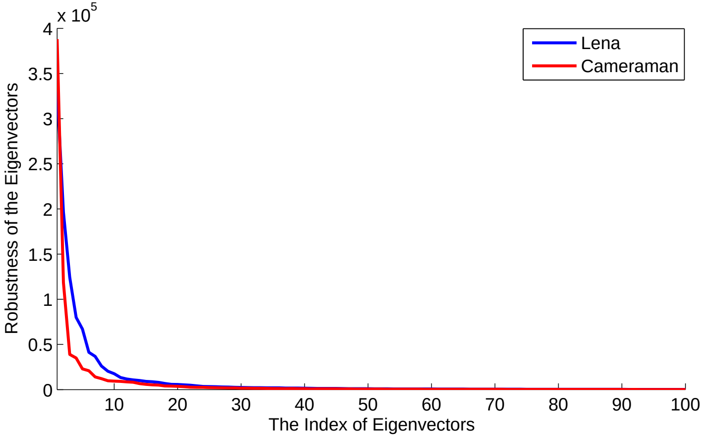


Figure: The robustness, i.e. total variance (see Equation (12)), of the first 100 eigenvectors using Lena and Cameraman images.

Table: Datasets descriptions.

Dataset	Number of Classes	Number of Features	Number of Samples
Iris	3	4	150
Iono	2	34	351
Ovarian	2	4000	216
ORL	5	10304	50
Ear$_{64\times 64}$	5	4096	30
Yale	5	77760	55

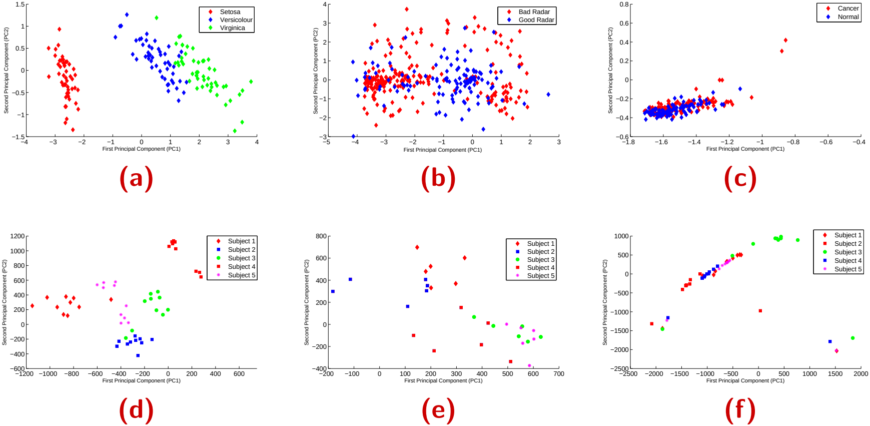


Figure: 2D Visualization of the datasets listed in Table 3, (a) Iris dataset, (b) Iono dataset, (c) Ovarian dataset, (d) ORL dataset, (e) Ear dataset, (f) Yale dataset.

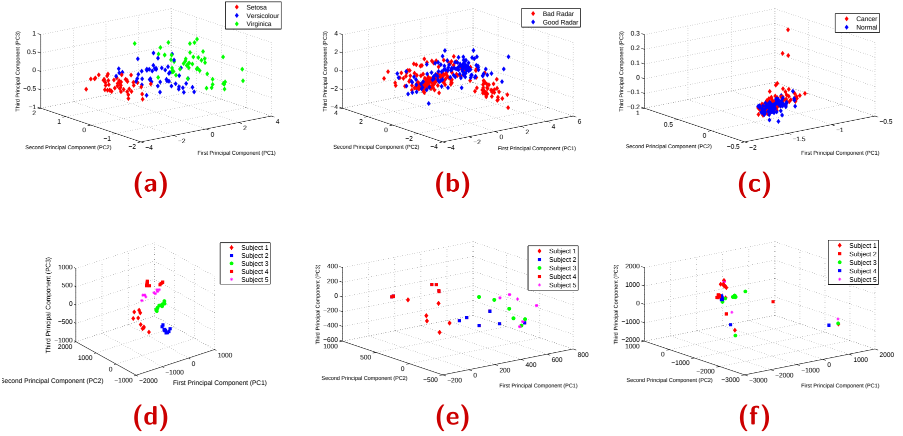


Figure: 3D Visualization of the datasets that were listed in Table 3, (a) Iris dataset, (b) Iono dataset, (c) Ovarian dataset, (d) ORL dataset, (e) Ear dataset, (f) Yale dataset.

Table: A comparison between 2D and 3D visualization in terms of MSE and robustness using the datasets that were listed in Table 3.

Dataset	2D		3D	
	Robustness (in %)	MSE	Robustness (in %)	MSE
Iris	97.76	0.12	99.48	0.05
Iono	43.62	0.25	51.09	0.23
Ovarian	98.75	0.04	99.11	0.03
ORL	34.05	24.03	41.64	22.16
Ear$_{64\times64}$	41.17	15.07	50.71	13.73
Yale	48.5	31.86	57.86	28.80

- How to calculate or construct PCA space.

For any other questions send to **engalaatharwat@hotmail.com**