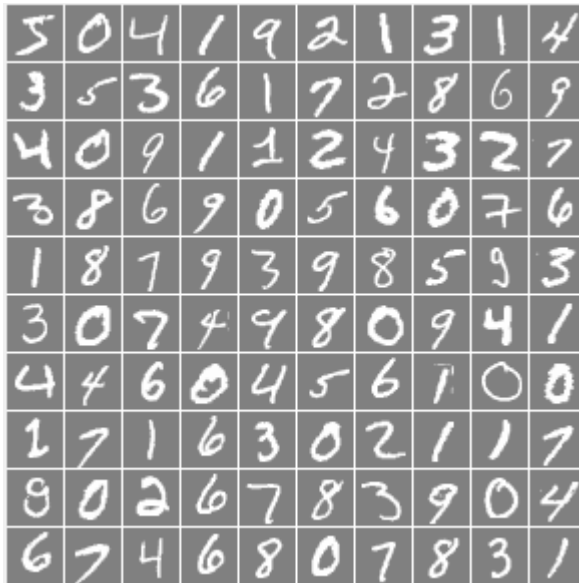


- (a) Load the first 1000 of the MNIST digits. Visualize the 100 first digits (using provided visual function) and check the corresponding class labels to ensure that the data format is correct (as a matrix: one row is one data object, column corresponds to data dimension). [3 points]

Jawaban :

```
[X y] = loadmnist( 1000 );  
visual(X(1:100,:));
```



- (b) Apply PCA on data as the following steps,

- i. Represents one data object as a column vector by transposing the matrix, thus the matrix now has 784 rows and 1000 columns. [3 points]

jawaban :

dikarenakan dalam matlab matrix kovarian membandingkan antar kolom (atribut) maka matrix tidak di transpose agar hasil kovarian 784 x 784.

- ii. Subtract the means of data thus the data now is zero-centered. [10 points]

```
rataRata = mean(Xi);
```

 rataRata 1x784 double


- iii. Compute covariance matrix C of the zero-centered data. [8 points]

```
CovMatrixC = cov(Xi);
```

 CovMatrixC 784x784 double

iv. Compute the eigenvectors v_i and eigenvalues λ_i of covariance matrix C . [9 points]

```
7 - [VectC NilaiC] = eig(CovMatrixC);
```

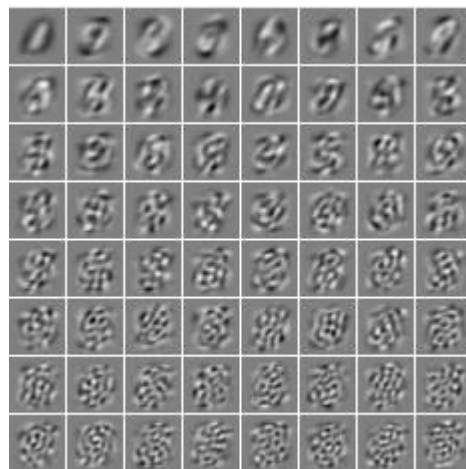
 NilaiC 784x784 double

 VectC 784x784 double

v. Sort descending the eigenvalues, and re-order the eigenvectors correspond to the order of eigen values. [10 points]

```
for i= 1 : 784
    nilaiCDesc = [ nilaiCDesc ; NilaiC(785-i,785-i)];
    a = []
    for j = 1 : 784
        a = [ a VectC(j,785-i)];
    end
    VectCDesc = [ VectCDesc ; a];
end
```

vi. Visualize the 64 first principal components (these are 64 eigen vectors correspond to the 64 first eigen values) using provided visual function. [3 points]



vii. Project the data into eigen subspaces (PCA domain) using 64 first principal components. [10 points]

```
Vrow = VectCDesc(1:64,:);
PCADomain = Vrow * B;
```

viii. What is the dimension of each data object in PCA domain now? [2 points]

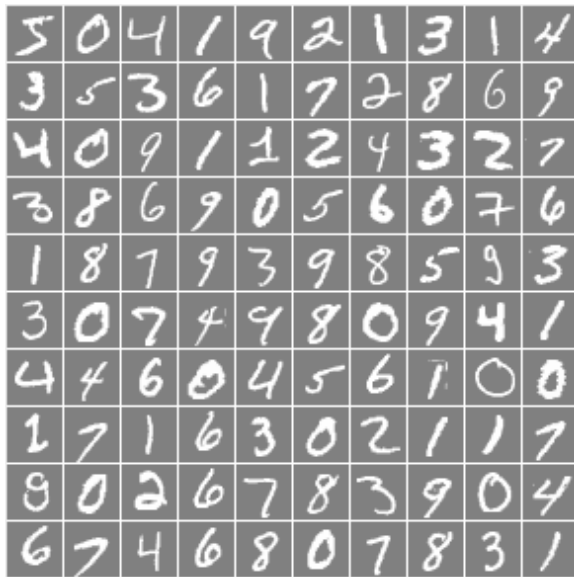
 PCADomain 64x1000 double

- (c) Reconstruct data from PCA domain (obtained from step vii) into native domain. Visualize 100 first reconstructed data and compare them to 100 first original data that are already visualized in (a). Give your opinion. [10 points]

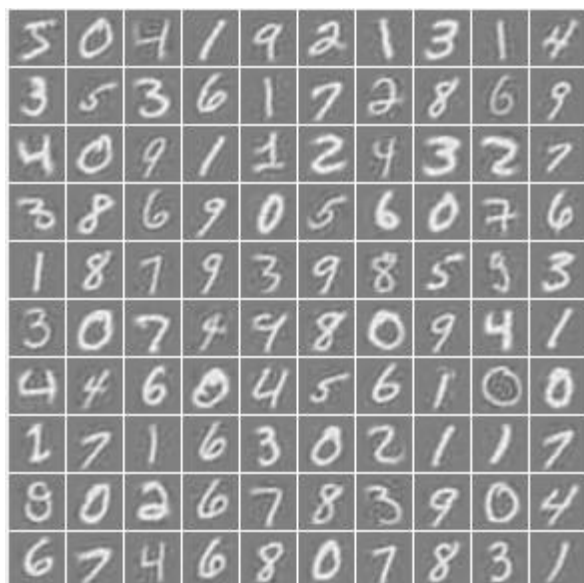
Jawaban :

```
DomainAsli = Vrow' * PCADomain;
visual(DomainAsli(1:100,:));
```

Visual awal



Visual rekontruksi dari domain PCA




Pendapat : hasil rekonstruksi dari domain PCA dengan 64 principal component cukup baik untuk mereduksi data, Karena ukuran matrix menjadi lebih kecil tetapi data yang kembali terrekonstruksi masih cukup baik.

- (d) What is the percentage of correct reconstruction obtained from (c)? [5 points] (Hints: the eigenvectors v_i of covariance matrix are the vector of each principal component, meanwhile the eigenvalues λ_i of covariance matrix are the variance of each principal component. The k first principal components explain the proportion of variance. Furthermore, the proportion of variance "explained" by PCA is the percentage of correct reconstruction.)

Jawaban :

```
a = 0;
for i=1 : 64
    a = a + nilaiCDesc(i);
end
b = 0;
for i=1 : 784
    b = b + nilaiCDesc(i);
end

nilaiHasilRekontruksi = (a/b) * 100;
```

 nilaiHasilRekontru... 87.9252

Nilai hasil rekontruksi dari domain PCA sebesar 87.9252%