HOMEWORK 2
DIMENSIONALITY REDUCTION

Vikas Jain 13788
Department of Computer Science and Engineering

Part A: Linear Dimensionality Reduction: Principal Components

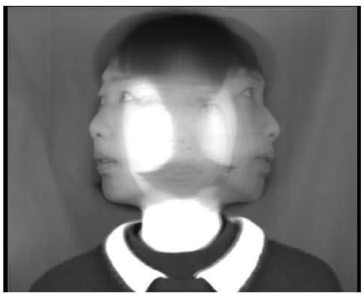
- An image set is given on which PCA has to be applied to obtain k-principal components. PCA allows to compute a **linear transformation** that maps data from a high dimensional space to a lower dimensional sub-space.

- **Algorithm used**
  1) n images of size d (576×720)
  2) finding average image *Imgavg*
  3) Subtracting mean image from all images and create a matrix X of all images. Size(X) is $d \times n$.
  4) Construct a covariance matrix $X^T X$ $(n \times n)$ and find its eigenvectors and eigenvalues and arrange them in decreasing order of eigenvalues. Eigenvector matrix of *k-principal component* is V $(n \times k)$.
  5) Original Eigenvector matrix is $U=X*V$ $(d \times k)$ (also normalized).
  6) Compress image into it weight, $w=U^T*(x-Imgavg)$ $(K \times 1)$, where x is the original image data $(d \times 1)$.
  7) Reconstruct image from weights $x^f=U*w+Imgavg$.

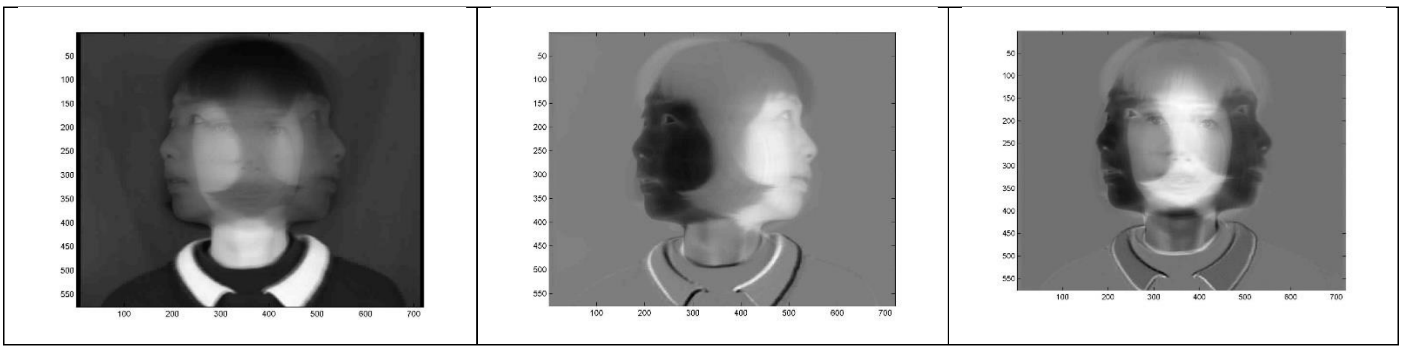
Link to dataset.
Link to Complete code.

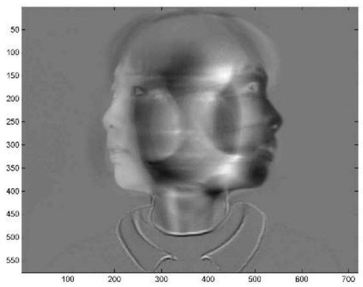
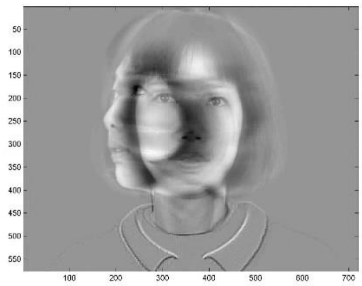
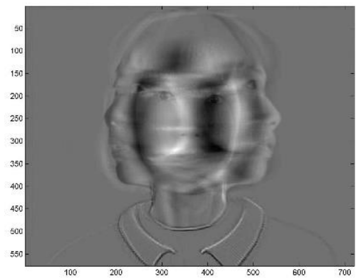
Observations:

No. of images, n=143; size of image, d= 576×720
Average face obtained:



Some Eigen faces Obtained from eigenvectors:

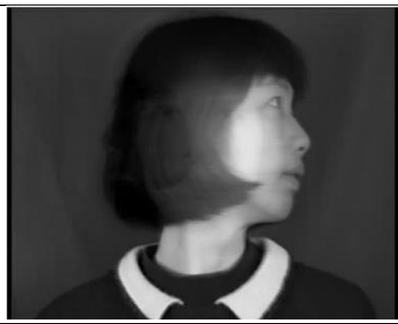


<u>Reconstruction of Image 25 from eigenvectors:</u>

Original Image 25:

Reconstructed images:

M=2	
M=10	
M=30	
M=80	

<u>DISCUSSION</u>.

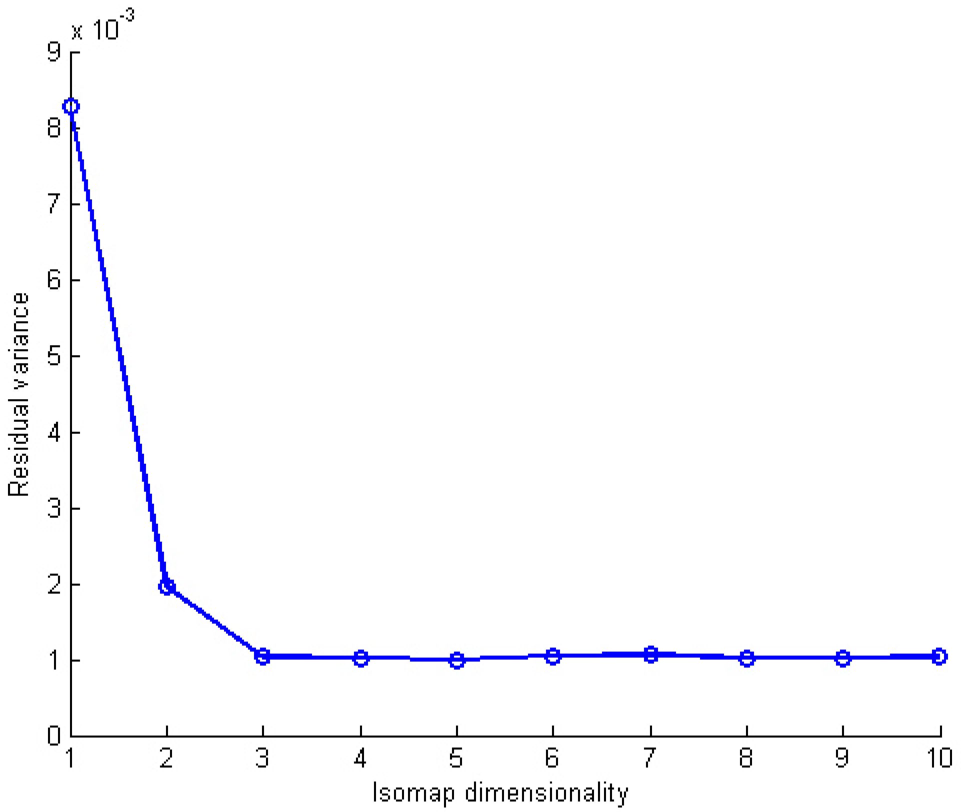
The PCA technique applied is able to reconstruct the images using linear combination of eigenvectors very well. It is also observable that if the eigenvectors used in the reconstruction are increased then the quality of image produced also increases. Also, since the original dimensionality in dataset is 1, the image reconstructed for m=2 is also acceptable.

PART B: Isomap

Isomap Algorithm is applied on the given data set to find the low dimension embedding of the data set. The observations of the Isomap algorithm are shown below.

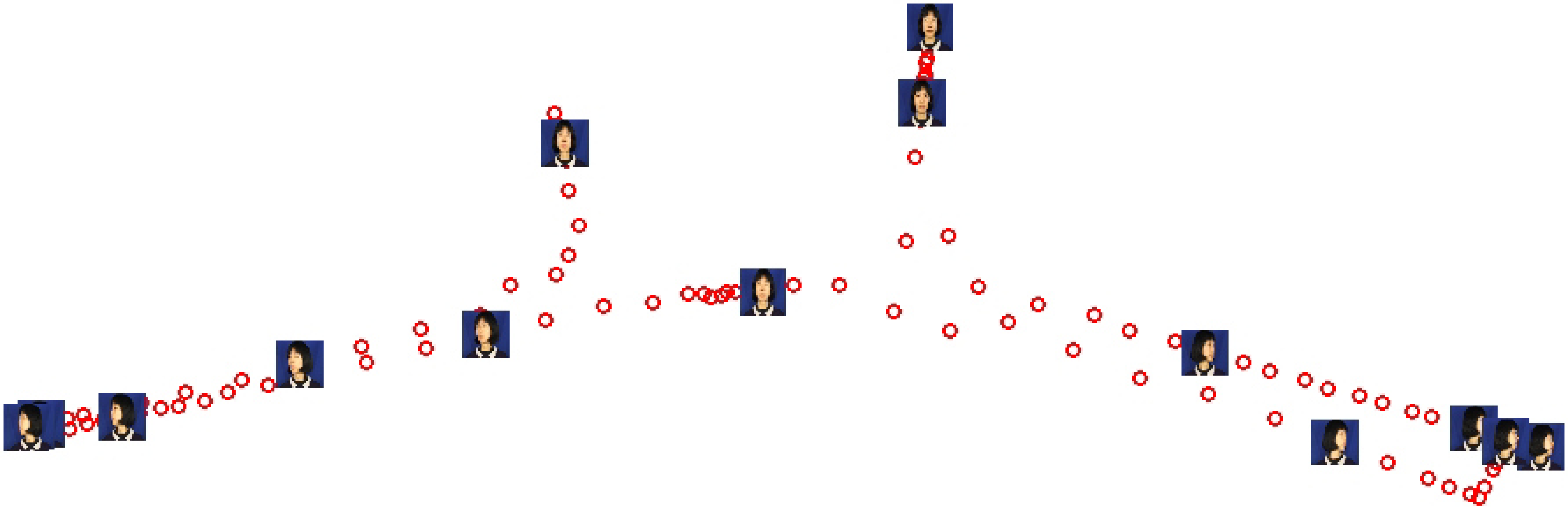
Observations

1) Residual Variance Plot for dimensions 1:10

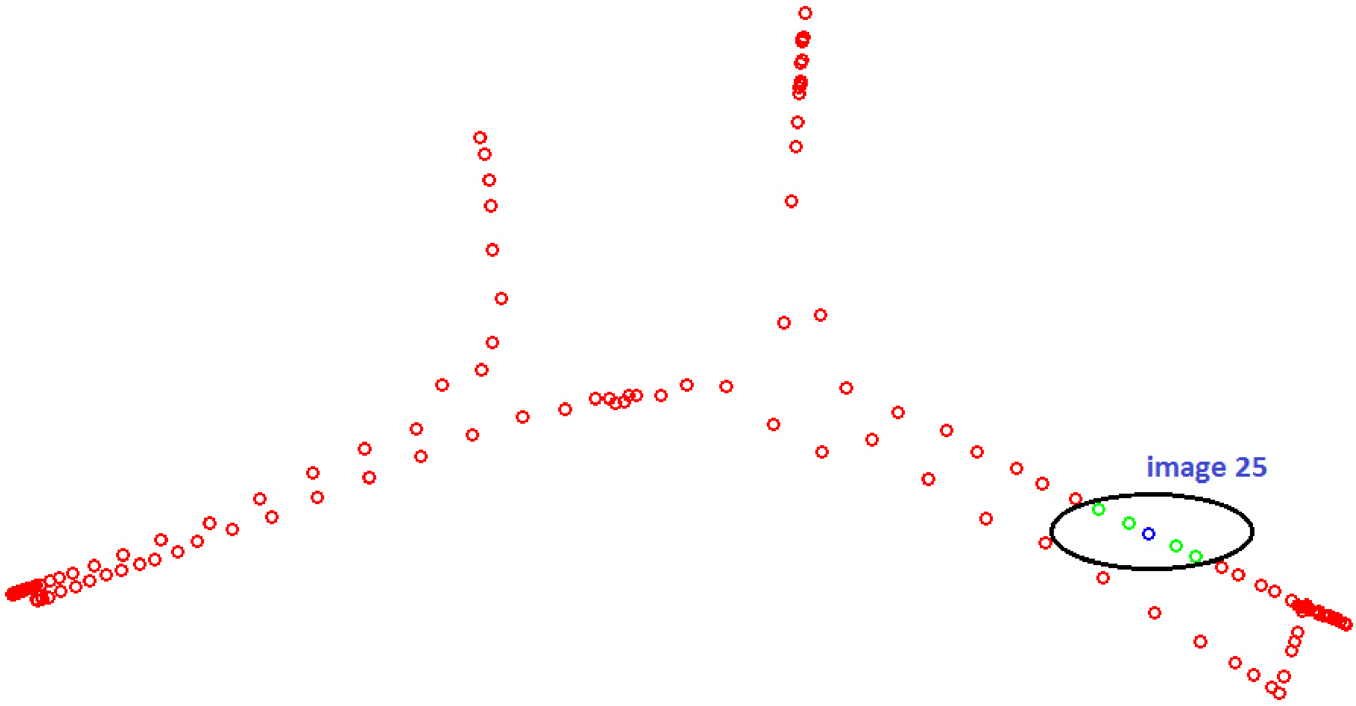


2) Isomap Embedding with Dimension = 2

Two-dimensional Isomap embedding (with neighborhood graph).



3) Image 25 neighbourhood

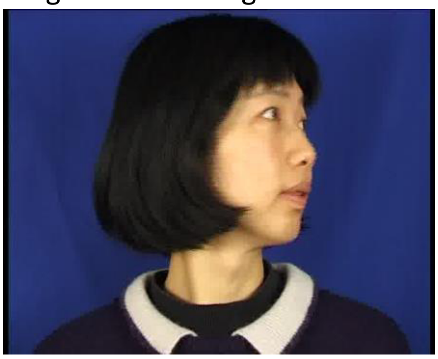


Blue circle: image 25

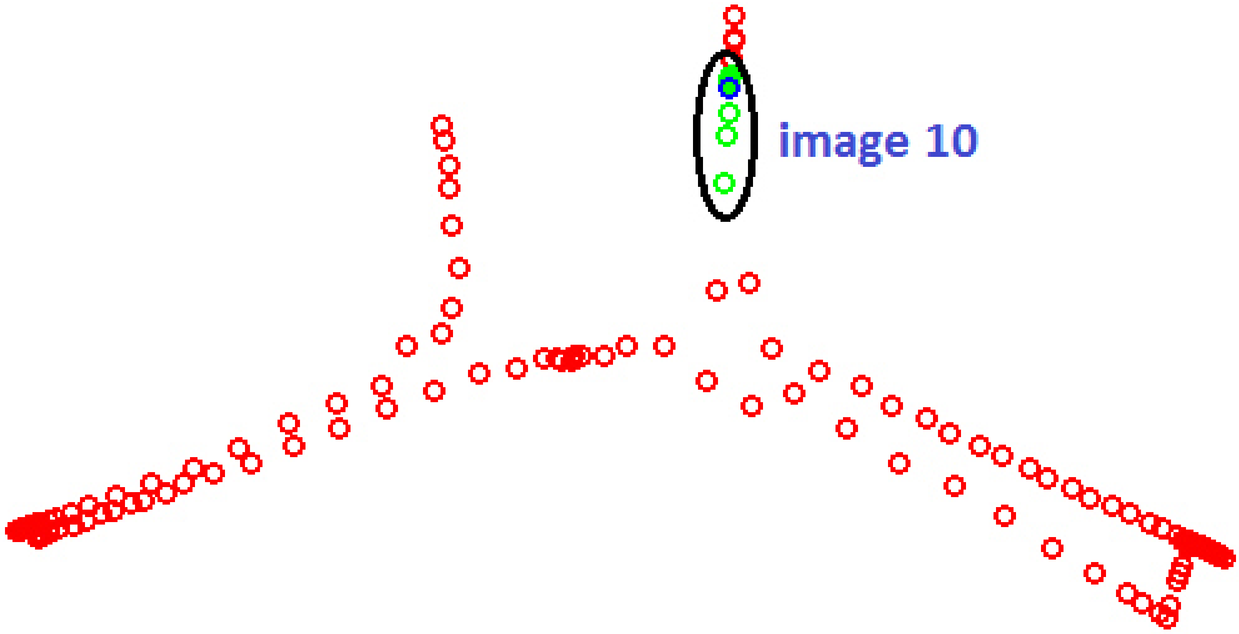
Green circle: neighbours of image 25

Image 25:

Neighbours of image 25:



4) <u>Image 10 in manifold</u>



Zoom image:

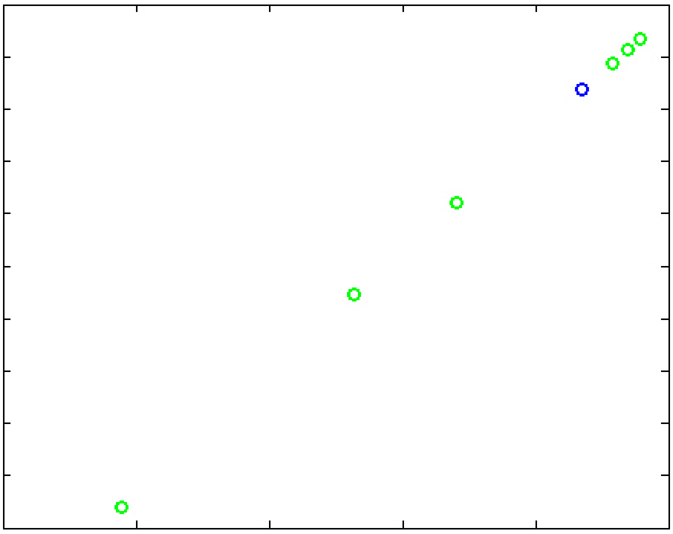


Image 10 (blue):

Blue Circle: Image 10; Green Circle: Neighbours

DISCUSSION

The dataset we have is only varying in one dimension because in all the images only head of the lady is rotating in one direction. But the data is highly non-linear and the Isomap applies NDLR method on the dataset. The manifold obtained from Isomap from low embedding dimension 2 is quite acceptable. It parts the data mainly into three categories (three linear segments in manifold) which are representing head in straight, right and left position.

The Residual variance plot suggesting that low dimension embedding of data should be 5 (least). But the value of residual variance is not varying much from 2 to 10.

Overall, Isomap is providing very good low dimension embedding for the given high dimensional data.

REFERENCES

PART A:

1) Principal component analysis. (2015, January 31). In *Wikipedia, The Free Encyclopedia*. Retrieved 20:13, February 9, 2015, from http://en.wikipedia.org/w/index.php?title=Principal_component_analysis&oldid=644950175

PART B:

1) V. d. S. Joshua B. Tenenbaum and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science vol. 290, pp. 2319{2323, December 2000.
2) Bhuwan Dhingra, " Local Quadrature Reconstruction on smooth manifolds", Masters Thesis, Electrical Engineering Department, Indian Institute of Technology Kanpur, July 2013.
3) Matlab Code Source : http://www.cse.iitk.ac.in/users/manifolds/theses/bhuwand/ by Bhuwan Dhingra