MATH 5473/CSIC5011 Project 2: Order the faces by Manifold Learning

Ng Yui Hong[1] yhngap@connect.ust.hk

[1]: Department of Mathematics, HKUST

1. Introduction

Face pose determination is an important area of research in human computer interaction (HCI). An important problem in HCI is to determine one's focus of attention. This can be inferred from the person's head orientation

In this project, we want to explore how manifold learning methods contribute to appearance-based methods so as to order head directions

1.1 Dataset

The dataset contains 33 faces of the same person (112×92×33) in different angles

2. Methods

1. MDS

MDS aims to recover Euclidean coordinate in given pairwise distance metric.

First compute D which from the Euclidean distance matrix between 2 data points. Then, compute B by $-\frac{1}{2}HDH^T$ where H is Housholder matrix. By eigen-decompistion of B $\bar{B} = U\Lambda U^T$ with $\Lambda = diag(\lambda_1, ..., \lambda_n)$, we can choose top k eigenvalues and eigenvectors to form the embedding data point $\tilde{T}_k = U_k\Lambda_k^{\frac{1}{2}}$ where $U_k = [u_1, ..., u_k]$, $\Lambda_k = diag(\lambda_1, ..., \lambda_k)$

2. ISOMAP

ISOMAP is a extended method of MDS in that it uses pairwise geodesic distances between data points and graph shortest path distances to reconstruct the data.

To be concise, the distance is $d_{ij} = min_{P=(x_i,...x_j)}(||x_i - x_{t_1}|| + ... + ||x_{t_{k-1}} - x_j||)$

3. Diffusion map

The embedding coordinates are computed by the eigen-decompision of a random walk matrix L on graph G=(V,E,W), where W is defined as $exp(-\frac{d(x_i,x_j)^2}{t})$

The random walk matrix L is defined as $L = D^{-1}W - I$ where $d_i = \sum_{i=1}^{n} W_{ij}$ and $D = diag(d_i)$

4. LLE/MLLE/HLLE

The central point of LLE is that any data point in a high dimensional space can be a linear combination of data points in its neighborhood.

Given a graph G. one can first do linear fitting $\min_{\sum_{j\in N_i} w_{ij}=1} \left\| x_i - \sum_{j\in\mathcal{N}_i} w_{ij}x_j \right\|^2$

Then, a global allignment $K = (I-W)^T(I-W)$ $W_{ij} = \begin{cases} w_{ij}, & j \in \mathcal{N}_i \\ 0, & \text{otherwise} \end{cases}$

Eigen decompose K thus computer the embedding like MDS

MLLE is another method to improve the performance of LLE. It uses multiple weight vectors projected from orthogonal complement of local PCA in each neighborhood

HLLE is also a method to solve the regularization problem of LLE. It revolves around a hessian-based quadratic form at each neighborhood.

5. LTSA

LTSA is a modified version of LLE.

Its first step is to do local PCA, i.e. computing local SVD on neighborhood of x_i

The second step is to do tangent space alighment, Finally, we do eigenvalue decomposition and find the smallest d + 1 eigenvectors of K with dropping the smallest one. The remaining d eigenvectors will give rise to d-mebedding coordinates

6. t-sne

3.1 Experiment

1. Re-arrange the image in order to form the ground truth
2. Run through each manifold learning method to see its prediction

Original#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Ground truth	8	13	19	32	6	18	28	7	17	1	5	16	12	10	4	21
Diffusion Map	9	11	19	31	5	18	28	6	17	1	7	16	12	10	4	21
MDS	9	10	19	32	5	18	30	6	17	1	7	16	12	11	4	21
ISOMAP	7	13	19	32	5	18	28	10	17	1	6	16	12	9	4	21
LLE	8	13	19	32	5	18	28	7	17	1	6	16	12	10	4	21
MLLE	8	12	19	32	6	18	28	7	17	1	5	16	11	10	4	21
HLLE	8	13	19	32	6	18	28	7	17	2	5	16	12	10	4	21
Spectral Embedding	6	13	18	32	8	19	28	7	17	3	5	16	12	10	4	21
LTSA	8	13	19	32	6	18	28	7	17	2	5	16	12	10	4	21
t-SNE	8	12	20	30	14	22	23	10	17	1	18	16	5	2	6	9

Original#	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
Ground truth	22	26	33	11	2	24	3	27	29	23	15	30	31	20	14	25	9
Diffusion Map	22	26	33	13	2	23	3	27	30	24	14	32	29	20	15	25	8
MDS	22	26	28	13	2	23	3	27	31	24	14	33	29	20	15	25	8
ISOMAP	22	26	33	11	2	24	3	27	29	23	14	30	31	20	15	25	8
LLE	22	26	33	11	2	23	3	27	29	24	14	30	31	20	15	25	9
MLLE	22	26	33	13	2	23	3	27	29	24	14	31	30	20	15	25	9
HLLE	22	26	33	11	1	23	3	27	29	24	14	31	30	20	15	25	9
Spectral Embedding	22	26	33	11	2	24	1	27	29	23	14	30	31	20	15	25	9
LTSA	22	26	33	11	1	23	3	27	29	24	14	31	30	20	15	25	9
t-SNE	33	25	24	27	15	11	4	32	29	21	19	28	31	13	7	26	3

Total absolute error which sum up absolute error is used as metrics

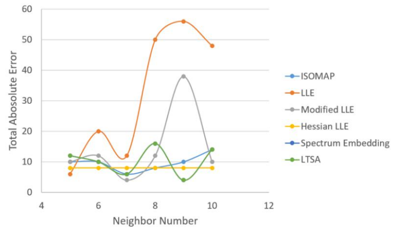
Methods	Diffusion Map	ISOMAP	LLE	MDS	HLLE	Spectral Emebedding	LTSA	MLLE	t-SNE
TAE	20	10	6	30	8	12	8	10	164

t-sne prform worse at face ordering task while LLE class done well.

3.Hyper-parameter study

Moreover, we explored the effect of number of neighbors (k) on the sorting performance for each method.

We tested for k =5 to 10, the methods were all applied with sklearn.manifold library. Compared with the other methods, LLE tends to be affected by k more. In terms of performance versus number of neighbors, we can find that 5 tends to be the best for most methods.



4. 2d embedding comparison

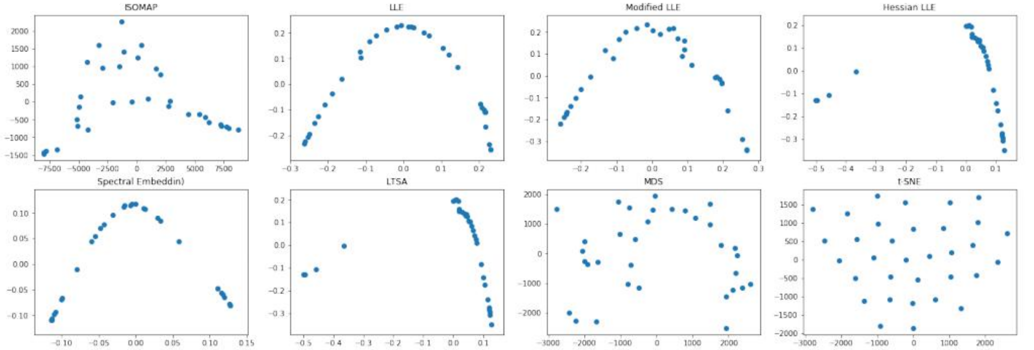
Afterwards, we compared the 2D embedding graph for different methods based on sklearn.manifold library. We assigned the number of neighbors to be 5 in our experiments.

We can see the result in next column.

Obviously, LLE and Spectral embedding seem to better than others in terms of x-axis which represents the first component. With the previous TAE result, we may know that LLE should be the best

3.2 Experiment(Cont.)

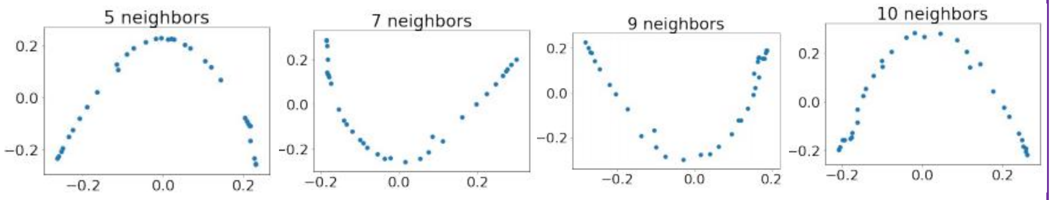
4. 2d embedding comparison(cont.)



One interesting phenomenon is that the shape of the graph can be both 'V' and 'Λ'.

However, this phenomenon makes no difference to the conclusion. The focus here is the discrepancy between different points in the figure

As the number of neighbors increases, more and more points tend to be indistinguishable. This provides us with another way to understand that the number of neighbors of 5 should behave the best.



4 Conclusion

In summary, almost all the methods we explored show reasonable sorting order except t-SNE. We quantified the sorting performance by calculating the total absolute error between different methods' results and ground truth and found that LLE exhibited the best performance. What's more, a parameter study was performed on the number of nearest neighbor k and 5 proved to be the best.

For future work, these methods can be applied to a more complicated dataset, (e.g. head images with both turning and nodding motion). By adding one degree of freedom, more eigenvectors may be involved in deciding the order. Last but not least, more techniques can be combined with the methods in this report to achieve emotion detection and so on

Reference

Youtube Presentation Link: https://youtu.be/jjX690khNNI

[1] Ji, Qiang. 3D face pose estimation and tracking from a monocular camera, Image and vision computing,

[2] G. Young and A. S. Householder, A note on multidimensional psycho-physical analysis, Psychometrika 6 (1941)

[3] David L. Donoho.Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for highdimensional data,