Information visualization by dimensionality reduction: a review

Safa A. Najim

Computer Science Department, Science College, Basrah University, Iraq
Email: safanajim@gmail.com

Abstract

Information visualization can be considered a process of transforming similarity relationships between data points to a geometric representation in order to see unseen information. High-dimensionality data sets are one of the main problems of information visualization. Dimensionality Reduction (DR) is therefore a useful strategy to project high-dimensional space onto low-dimensional space, which it can be visualized directly. The application of this technique has several benefits. First, DR can minimize the amount of storage needed by reducing the size of the data sets. Second, it helps to understand the data sets by discarding any irrelevant features, and to focus on the main important features. DR can enable the discovery of rich information, which assists the task of data analysis. Visualization of high-dimensional data sets is widely used in many fields, such as remote sensing imagery, biology, computer vision, and computer graphics. The visualization is a simple way to understand the high-dimensional space because the relationship between original data points is incomprehensible. A large number of DR methods which attempt to minimize the loss of original information. This paper discuss and analys some DR methods to support the idea of dimensionality reduction to get trustworthy visualization.

Keywords: *Dimensionality Reduction, Information visualization, Information retrieval.*

1. Introduction

Visualisation of high-dimensional data sets is widely used to analyse data in many fields of study, including remote sensing imagery, biology, computer vision, and computer graphics. Its purpose is to provide rich information to assist with data analysis [1]. Dimensionality reduction (DR) is an important step for data pre-processing in visualisation and knowledge discovery, and it is used for different purposes, such as information visualisation, noise reduction, and imaging applications [2] [3]. Formally, for a set of n input points $X \subset \mathcal{R}^D$, $\phi(X)$ is used to project the D-dimensional data points $x_i \in X$ to d-dimensional data points $y_i \in Y$, where $d \ll D$.

$$\phi : \mathcal{R}^D \rightarrow \mathcal{R}^d \tag{1}$$
$$x_i \mapsto y_i \ \ \forall \ 1 \leq i \leq n \tag{2}$$

In Equation 1, ϕ attempts to approximate the output pairwise distance $d(x_i, x_j)$ with their corresponding in input space $r(y_i, y_j)$, i.e, $r(x_i, x_j) \approx d(y_i, y_j) \ \ \forall \ 1 \leq i \leq n$ to project X's data point correctly in Y space.

The high-dimensional data sets have several features; however, some might not be relevant to specific data analysis. DR is used to discover the main and important features by which to make analysis and visualisation possible. The fundamental information in the original data sets is reflected in the distances between pairs of data
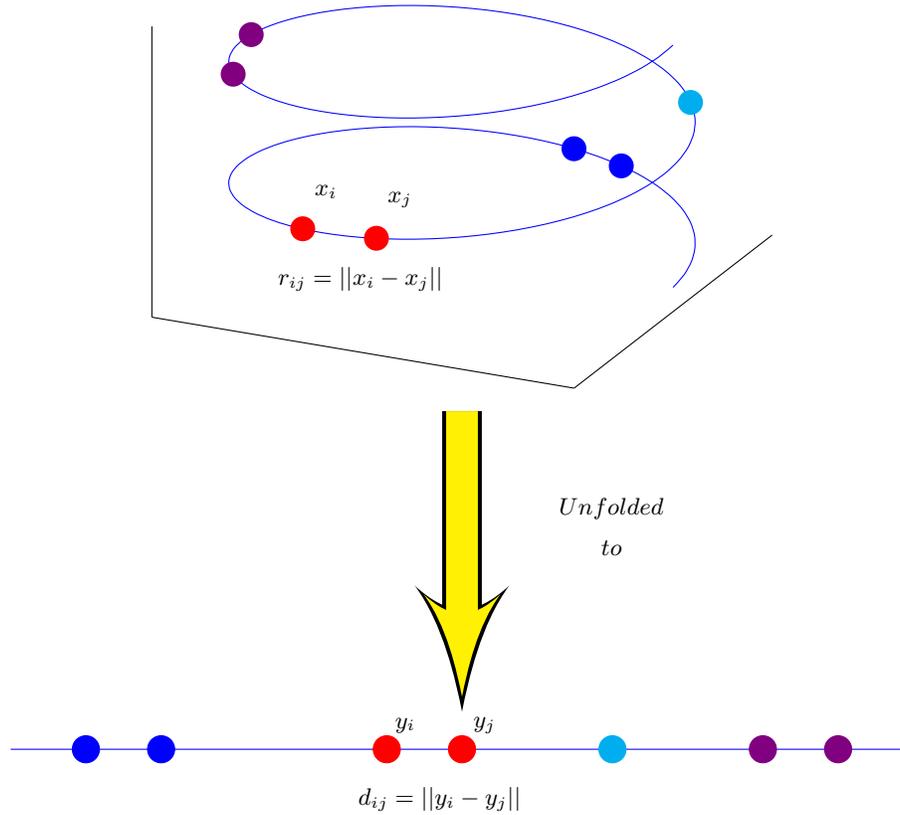
Figure 1: *Three-dimensional spiral data sets are unfolded to a one-dimensional straight string line. The neighbourhood relations between points in the unfolding space are preserved with their correspoinding relations in the high-dimensional data sets.*

points, and this information should be preserved by using a gradient step and fitting the input distances r_{ij} to output distances d_{ij}. Thus, the goal of preserving the distance is to represent the original data sets in a projectec space [4]. A simple example of applying DR is shown in Fig.1, where the spiral data sets are unfolded to the straight string. Each point in the straight string preserves its local neighbourhood relations, and preserving neighbourhood relations between points indicates the efficiency of the low-dimensional space.

In reality, reducing dimensionality of large data sets to a low-dimensional space without losing information might be impossible. In general, DR attempts to minimize as in the following equation:

$$\phi = \sqrt{\sum_{i,j=1}^{n} (r_{ij} - d_{ij})^2} \tag{3}$$

where $r_{ij} = ||x_i - x_j||$ and $d_{ij} = ||y_i - y_j||$ for the four points x_i, x_j, y_i and y_j. The cost function in Equation 3 measures the difference between the distances in the input space and the corresponding distances in the projected space, and the final values should be minimized according to the data in a projected space.

Unfolding a complex high-dimensional data sets into low-dimensional representation should focus on preserving the nearby neighbourhood relationship between points rather than on creating additional points. There are two ways by which to define neighbouring points for a point. The first supposes that all points are neighbours for a point, but the nearest k points are strong neighbours [5]. Each point has a fixed number of neighbours and this number will not change through the projection process. The second method uses a fixed circular radius r_c, where the neighbouring points are inside this domain for a point [6]. Thus, the number of neighbours is not the same for all points in the space.

2. Types of Dimensionality Reduction

A variety of strategies has resulted in the development of many different DR methods, which can be classified as being either linear or nonlinear. Most of the linear methods are non-iterative, which assumes that the data are distributed close to a hyperplane of original space to reflect their uniqueness [2]; therefore, they are constrained in many applications. To overcome this problem, most nonlinear methods adopt iterative optimisation to get more flexibility in searching for a representation of data points in a projected space. Iterative methods start with a random configuration of low-dimensional space, and iteratively refine it until the error measure of low-dimensional space is minimised to the defined threshold or the specified number of iterations has been sapped. Local and trustworthiness methods are two other types of DR.

Preservation of high-dimensional data sets distances depends on the distance similarity measure. Distance, in nature, is used to measure the dissimilarity between two data points in a space. Euclidean and geodesic distances are usually used in this matter. The notion of Euclidean distance was originally defined as the distance of a straight line between two points, but the efficiency of Euclidean distance depends on the type of area. For example, the distance between two nearby points is computed exactly by Euclidean distance; however, in nonlinear space, Euclidean distance cannot measure the correct distance. For this, the concept of geodesic distance is suitable for computing the shortest path between two points that lie on a surface, as in Fig.2. Although both ideas are used for distance preserving DR, many methods are enhanced when geodesic distance is used because it provides a better measurement of distance between pairs of data points than Euclidean distance.
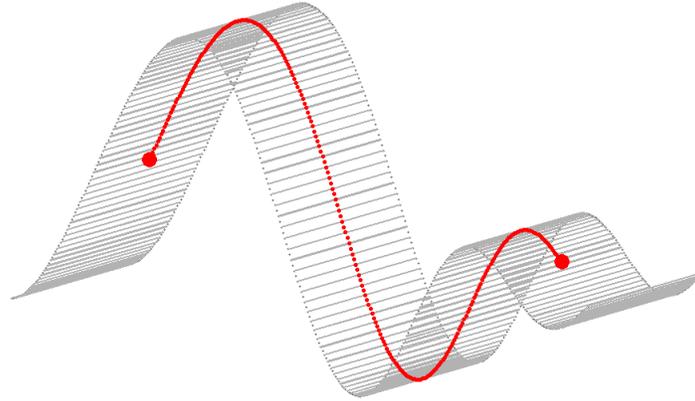

Figure 2: *Geodesic distance idea. The geodesic distance between the two red points is the length of the geodesic path, which is the shortest path between the points that lie on the surface.*

2.1. Linear Dimension Reduction Methods

Principle component analysis (PCA) is the oldest of the three methods, and has been used in different applications because of its simplicity. PCA supposes the linear projection between original data and projected space [7], but this principle does not work well with nonlinear data sets. The neighbourhood relationships between points are lost in projected space and do not preserve their nonlinear relationships of original data.

Fig.3(b) shows PCA cannot unfold Sphere data sets (in Fig.3(a)). PCA uses orthogonal linear combination to find linear transformation space of data sets $X \subset \mathcal{R}^D$. The steps of applying it are:

1. Compute the centroid of original data sets:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{4}$$

and subtract it from x_i to generate new origin

$$z_i = x_i - \mu \tag{5}$$

2. Compute the covariance matrix

$$cov = \frac{1}{n} \sum_{i=1}^{n} z_i \, z_i^T \tag{6}$$

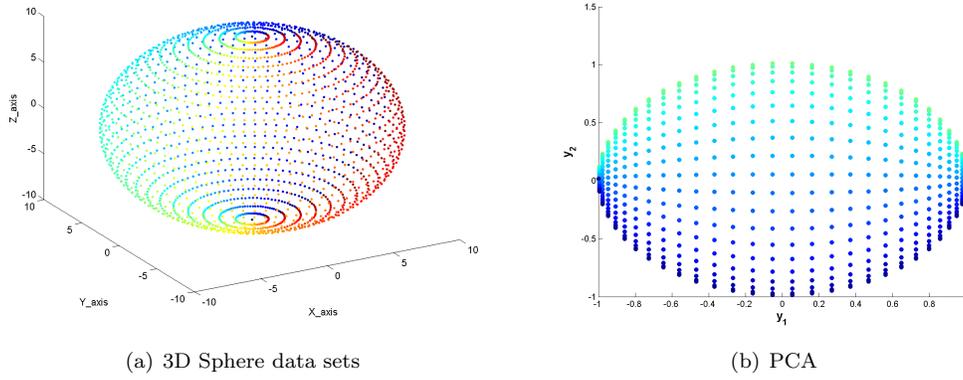(a) 3D Sphere data sets   (b) PCA

Figure 3: *a) Three-dimension Sphere data sets. b) Linear DR method (PCA) cannot unfold the Sphere data sets in two dimension representation, where overlap in neighbourhood relation is the predominate status of these unfolded representations.*

where z_i^T is the transpose of the z_i

3. According to the target dimension of low dimension space, PCA computes eigenvectors $u_1, u_2, ..., u_d$ and eigenvalues $\lambda_1, \lambda_2, ..., \lambda_d$ of covariance matrix cov.

4. Compute principal component transformation :

$$y = P^T z \tag{7}$$

where $P = [u_1, u_2, ..., u_d]$ and its transpose is P^T.

2.2.   Nonlinear Dimension Reduction Methods

Isometric feature mapping (Isomap) is a well-known global nonlinear DR method. Isomap attempts to preserve and hold the global structure of the original high-dimensional data sets [8] by computing the geodesic distance $r_{ij} = ||x_i - x_j||$ between two high-dimensional pairs of data points x_i and x_j. Geodesic distance is better than Euclidean distance in recognizing distances between neighbours near and far. Isomap is a continuity method that attempts to preserve the geodesic distance r_{ij} in the original space with its corresponding linear distance d_{ij} in the projected space. The algorithm of Isomap can be summarized in the following steps:

1. If the data sets are very large, select n random data points. Otherwise, all data points are selected.

2. Constructing the neighbourhood graph by connecting each data points to its k nearest neighnours.

3. The shortest paths among data points are computed by using Dijkstra's algorithm [9] to construct a geodesic distance matrix.

4. Applying multidimensional scaling (MDS) method to find low-dimensional space by minimizing the following stress function:

$$\phi(Y) = \sqrt{\sum_{i<j} (r_{ij} - d_{ij})^2} \tag{8}$$

5. If step 1 is applied, the interpolation is applied on the remaining data points to generate final projected space.

Isomap has been used in different applications by reducing the high-dimensional data sets into two or three dimensions [10]. It has the ability to deal with nonlinear data sets and can discover the information and details that were hidden on MDS. However, it cannot unfold the closed data sets, as Sphere; for example, Fig.4(a) shows an Isomap cannot unfolded Sphere data sets in a two-dimensional space. In addition, an extensive time commitment and storage requirements make Isomap inappropriate for use on large data sets.
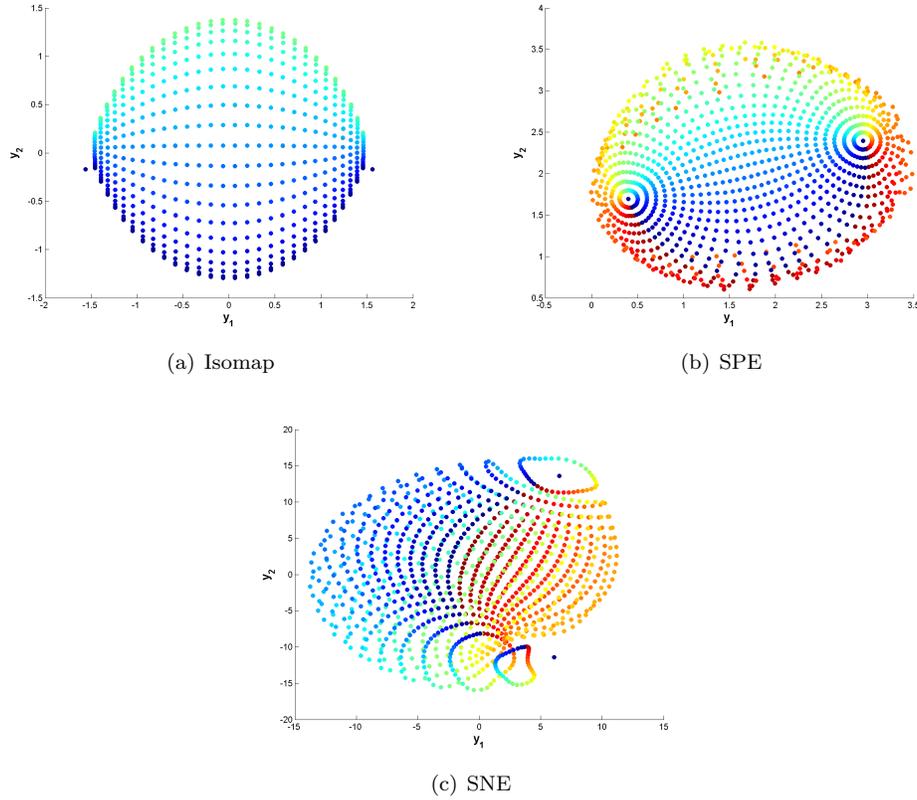
(a) Isomap

(b) SPE

(c) SNE

Figure 4: *Isomap and SNE cannot unfold the Sphere data sets (in Fig.3(a)). SPE unfolds it, but the visualization are projected with false neighbourhood points.*

Stochastic proximity embedding (SPE) is a nonlinear method that proceeds by calculating Euclidean distance for global neighbourhood points within a fixed radius [6]. SPE is an enormous step in computational efficiency over MDS, and is faster than Isomap. SPE is used in different applications and has succeeded in getting satisfactory results. The objective of SPE is to find representation that has points distances that are identical to their corresponding distances in high-dimensional data sets. The method starts by selecting a random point from original data, in time t, to be projected in the low-dimensional space.

Projected space starts with initial coordinates, and is updated iteratively by placing y_i and y_j onto the projected space in such a way that their Euclidean distance ($d_{ij} = ||y_i - y_j||$) is close to the corresponding distance ($r_{ij} = ||x_i - x_j||$) in original high-dimensional data sets. Thus, SPE minimizes the following Equation:

$$Stress = \sqrt{\frac{\sum_{i<j}(d_{ij} - r_{ij})^2 / r_{ij}}{\sum_{i<j} r_{ij}}}. \tag{9}$$

The points in projected space are updated according to the following constraint:

$$if \ (r_{ij} \leq r_c) \ or \ ((r_{ij} > r_c) \ and \ (d_{ij} < r_{ij}))$$

$$y_i \leftarrow y_i + \lambda(t)\frac{r_{ij} - d_{ij}}{d_{ij} + \epsilon} \ (y_i - y_j) \tag{10}$$

$$y_j \leftarrow y_j + \lambda(t)\frac{r_{ij} - d_{ij}}{d_{ij} + \epsilon} \ (y_j - y_i)$$

where $\lambda(t)$ is learning rate at t time, and r_c is a fixed circular radius of neighbourhood points. Fig.4(b) showed how the SPE unfolds the Sphere data sets, where the distances among projected space points are preserved with their corresponding distances in the original data sets. However, the visualization contains false neighbourhood points.

The difficulty with using SPE is in determining the value of r_c. The results will be torn and very bad if r_c is very small, and SPE will be equivalent to MDS when r_c is very large.

The stochastic neighbour embedding (SNE) method is an iterative nonlinear method that attempts to preserve structural properties between low-dimensional space pairwise data points with their corresponding distance in high-dimensional data sets. It computes asymmetric probability p_{ij} between two neighbour points x_i and x_j in high-dimensional data sets.

$$p_{ij} = \frac{e^{-||x_i - x_j||^2/2\sigma_i^2}}{\sum_{k \neq i} e^{-||x_i - x_k||^2/2\sigma_i^2}} \tag{11}$$

where $p_{ii} = 0$, and σ_i is variance of Gaussian centered around x_i set by user. Asymmetric probability q_{ij} of the corresponding points in low-dimensional space y_i and y_j of original high-dimensional points x_i and x_j are computed by:

$$q_{ij} = \frac{e^{-||x_i - x_j||^2}}{\sum_{k \neq i} e^{-||x_i - x_k||^2}} \tag{12}$$

where $q_{ii} = 0$. SNE attempts to find low-dimensional space by matching p_{ij} and q_{ij} as much as possible by minimizing stress function in Equation 13 instead of using squared differences between p_{ij} and q_{ij}:

$$S = \sum_i \sum_j p_{ij} \ log \ \frac{p_{ij}}{q_{ij}} \tag{13}$$

and S can be minimizing by using a gradient descent method. With closed data sets, as Sphere, the SNE cannot unfold it, as in Fig.4(c). Many of the points are overlapped with others that causes to loss some information.

2.3.   Local Nonlinear Methods

Locally linear embedding (LLE) is a nonlinear dimension reduction method that attempts to project the nearby neighbourhood points to a locally linear projected space. LLE uses three steps to do this task; first, it finds k nearest neighbourhood points for each point in high-dimensional data sets. Second, it computes a weight matrix among neighbours W by minimizing Equation 14. Weights among neighbours represent a value of strength relation among them.

$$\varepsilon(W) = \sum_i^n \|x_i - \sum_{j \neq i} w_{ij} x_j\|^2 \tag{14}$$

where $w_{ij} = 0$, if $x_j \neq x_i$. $\forall i, \sum_j w_{ij} = 1$. In the final step, LLE lies the D-dimensions of original data into d-dimensions space, where $d \ll D$. It uses weight matrix W to find the coordinates representation of low-dimensional space in order to preserve the topology properties of the original data. Based on locally linear reconstruction, LLE minimizes the following equation to find the projected space

$$\Phi(Y) = \sum_i^n \|y_i - \sum_{j \neq i} w_{ij} y_j\|^2 \tag{15}$$

subject to $\frac{1}{N} \sum_i y_i = 0$, and $\frac{1}{N} Y^T Y = I$. In many applications, LLE has poor generalization because it is not dynamic [11]. As an example, Fig.5 shows that LLE cannot unfold the three dimension of the Sphere data sets. it forces projected space to have a representation according to the weight matrix of high-dimensional data sets. Thus, LLE is poor in dealing with more complex data sets.

2.4.   Trustworthiness Methods

The curvilinear component analysis (CCA) method preserves the pairwise distances in the low-dimensional space with their corresponding pairwise distance in the original high-dimensional data sets by minimizing the following cost function:

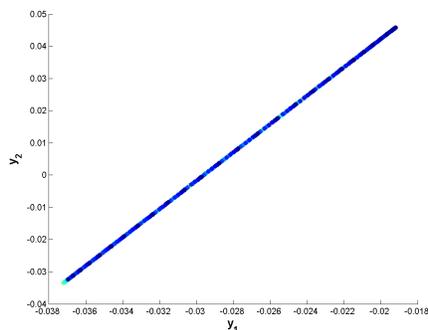$$\phi(Y) = \sum_{i<j} (r_{ij} - d_{ij})^2 F(d_{ij}, \lambda_t) \tag{16}$$

Figure 5: *LLE method cannot unfold the Sphere data sets, in Fig.3(a), in a satisfactory low-dimensional manifold.*

where $r_{ij} = ||x_i - x_j||$ and $d_{ij} = ||y_i - y_j||$ are the Euclidean distances between data points i and j in original high-dimensional and low-dimensional spaces, respectively. F is a bounded decreasing function, and allows CCA to preserve the distances on different scales depending on the time dependent value of λ_t which is started with large value to cover all data points, and then gradually decreased throughout processing. It is defined as:
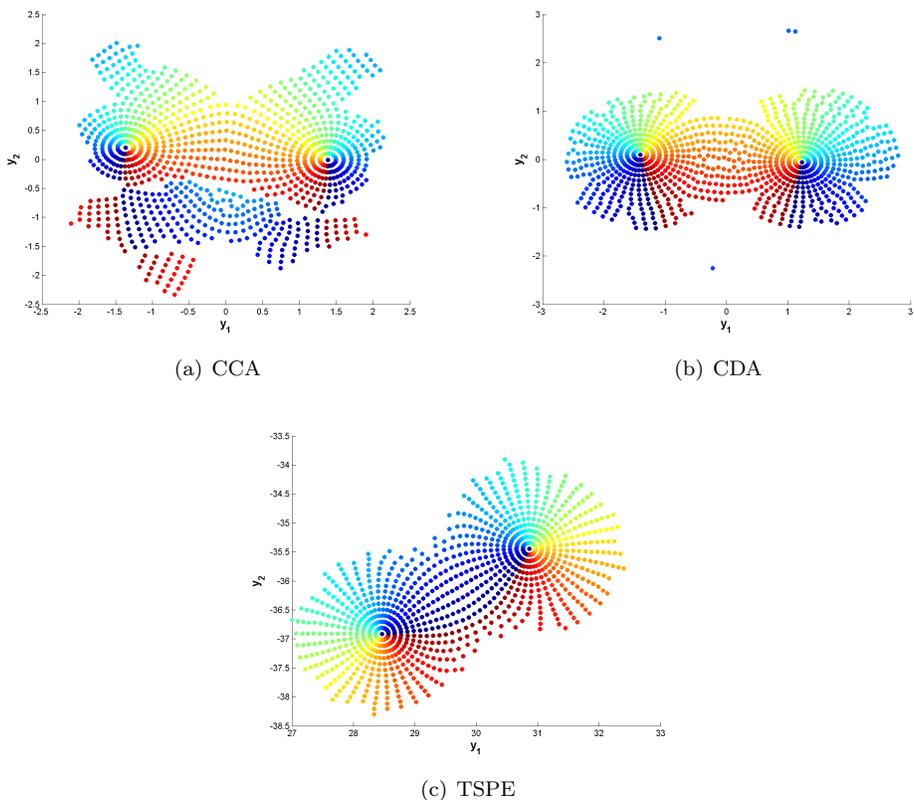

(a) CCA

(b) CDA

(c) TSPE

Figure 6: *Trustworthiness methods, CCA, CDA and TSPE, can unfold the Sphere data sets. TSPE is better than CCA and CDA because the original information of the Sphere data sets is preserved in the low-dimensional manifold.*

$$F(d_{ij}, \lambda_t) = \begin{cases} 1, & \text{if } d_{ij} \leq \lambda_t \\ 0, & Otherwise \end{cases} \tag{17}$$

CCA can find the projected space of some nonlinear data sets, as in the unfolding Sphere data sets in Fig.6(a). However, it fails with many nonlinear data sets. Thus, preserving the neighbours distance is not guaranteed by CCA.

The curvilinear distance analysis (CDA) method is a version of CCA, and attempts to preserve the pairwise distances in the low-dimensional space with their corresponding pairwise distance in the original high-dimensional data sets [12]. Like Isomap, CDA uses geodesic distance to keep the preservation between two distances. CDA uses the same CCA Equations in 16 and 17, but r_{ij} is the geodesic distance between data points i and j in the original high-dimensional data sets, and d_{ij} is the Euclidean distance between data points in the low-dimensional space. The general steps of applying CDA are:

1. If the data sets are very large, select n random data points. Otherwise, all data points are selected.

2. Constructing the neighbourhood graph by connecting each data-point to its k nearest neighnours.

3. The shortest paths among data points are computed by using Dijkstra's algorithm [9] to construct geodesic distance matrix.

4. Applying CCA algorithm using processed geodesic distances.

5. If step 1 is applied, the interpolation is applied on the remaining data points to generate final projected space.

CDA can find unfolded space of traditional nonlinear data sets, such as Sphere in Fig.6(b). The preserving original information by CDA is better than CCA, especially with complex data sets. It prevents points in the projected space from overlapping [13]. Thus, geodesic distance with the efficiency of CCA gives the CDA more performance. However, the low-dimensional manifold might be torn when the neighbourhood relation of the original high-dimensional data sets are lost.

Trustworthy stochastic proximity embedding (*TSPE*) has the ability to visualize difficult data sets, and, in terms of visualization, it gives satisfactory results [14]. This ability comes from dealing adequately with the projected space. In general, a projected space is improved through projection process, consequently, *TSPE* focus on this state by using decreasing neighbourhood size in order to continue this improvement. In each step of the projection process, the neighbourhood size is reduced in order to keep pace with improvements in the projected space. In addition, the space optimization is reduced gradually, with *TSPE* focusing, to begin with, on maintaining a distant relationship between the points and then maintaining the nearby relationships. The points in projected space are updated depending on their relation, as follows:

$$y_j \leftarrow y_j + \lambda(t) \ T(d_{ij}) \ \frac{r_{ij} - d_{ij}}{d_{ij} + \epsilon} \ (y_j - y_i) \tag{18}$$

$$T(d_{ij}) = \begin{cases} 1 & if \ (d_{ij} \ \leq \ d_c(t)) \ \vee \ ((d_{ij} > d_c(t)) \ \wedge \ (d_{ij} < r_{ij})) \\ 0 & Otherwise \end{cases} \tag{19}$$

where $d_c(t)$ is a decreased neighbourhood radius over time. *TSPE* starts iteratively, on the projected space, with selecting a random point in time t, which updates all the local neighbourhood points in a sufficient region within local neighbourhood radius $d_c(t)$, so the coherent structure will be constructed by sending false neighbours away; according to Equation 18. The local neighbourhood radius d_c starts with large value at t_0 to include all points, and then gradually this value is decreased through times by $d_c(t_0)/(t + 1)$ to keep neighbourhood points with the improvement of projected space. *TSPE* uses projected space and a decreasing neighbourhood radius $d_c(t)$ in the definition of $T(d_{ij})$, in Equation 18, which makes the proposed method overcomes the DR problems. Fig.7 shows the general idea of the *TSPE*. TSPE can find unfolded space of the Sphere data sets, as in Fig.6(c). The manifold dose not have false neighbourhood, and all the original information is preserved in the unfolding space.

3. Information Visualization

Information visualization is a way to present the information in a way that allows people to see things that were unseen before. It makes the information easy to share among the people by asking each other, "Do you see this thing?" In addition, it leads to more enhancement, such as, "If we made an enhancement to that area, what would happen?" Recently, information visualization has become the art of science, where there are many methods that can be used to help the people understand the meaning of the data by observing the proximity relationship among data points in a projected space. The colour of each pixel in the visualization is a compendium of information in the original data in which their most salient features are captured [15].
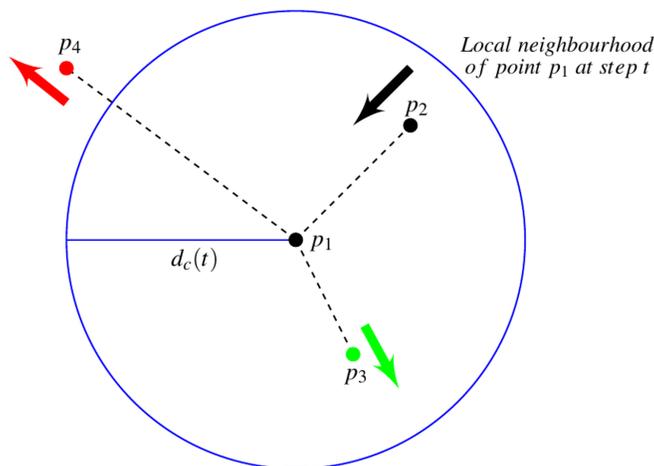
Figure 7: *The main idea of TSPE. At step t of iteration process, the point $p_1 \in Y$ is selected. The radius of local neighbourhood of this point at this step is $d_c(t)$. p_1 preserves its distance with true neighbours, as with p_2, when $(d_{ij} < d_c(t)) \wedge (d_{ij} \geq r_{ij})$. It pushes away the false neighbour points, as with p_3, when $(d_{ij} < d_c(t)) \wedge (d_{ij} < r_{ij})$. The points which are outside the local neighbourhood, which their $((d_{ij} > d_c(t)) \wedge (d_{ij} < r_{ij}))$, as with p_4, are pushed further away.*

Modern visualization tools have improved our ability to study many things directly from data. While there are different types of data, the way in which we visualize them are also different [16]. Visualizing scalar data is a popular way that is used in different sciences, where colours are mapped to the points. The visualization uses the same geometry as the original scale data, which can be 1D, 2D or 3D. The main advantage of this method is its simplicity and its ease of understanding. The visualization of vector field data sets has become more interesting in recent years. Vector data is scalar data with description. For example, to visualize fluid data sets, arrows, streamlines or animation are used to represent the direction. The visualization of vector field data sets is based on the dimensionality of the data sets. Three-dimensional vector data sets are a much more challenging problem, and the challenge increases when adding more descriptions (dimensions). The reconstruction of a three-dimensional volume model from a sequence of two-dimensional image slices of the human body is an example of volume rendering, which is helps in planning treatment or surgery.

Information visualization faces many challenge. For example, the visualization of flow data is difficult. Its visualization requires us to use different things to imitate the fluidity of the flow. In addition, the visualization of a connected network requires us to represent the relationships between nodes. However, finding a good representation containing all the information is a challenge. Many visualizations of data sets are pretty, but they do not show the important information. More powerful tools are required to meet the users needs. For example, the parallel coordinates method is used to represent high-dimensional data points into 2-dimensional space by using parallel axes. Although this method allows us to view the useful information, it is limited when it comes to representing large data sets. Information visualization can be used for different purposes:

1. Explanation: The data sets are visualized to explain something. For example, it is used to convince the viewer that one solution is better than another.

2. Exploration: The visualization of scientific data sets helps the researcher to explore the relationship between data sets.

3. Expression: Visualization is a way to represent data sets in a pretty way without giving more detail, in that we are focusing on the aesthetics.

The DR method is a strategy used to visualize a high-dimensional data sets by projecting a high-dimensional data set onto a low-dimensional space where it can be visualized directly [17]. The problem of DR methods arises in discovering the low-dimensional space in the high-dimensional complex data sets. In general, there are two types of DR methods: continuity and trustworthiness. Continuity methods, such as Isomap and SPE, attempt to inherit the relations among global neighbourhood points to local corresponding points in projected space. Although this procedure has advantages in strengthening the coherence of local corresponding points, continuity errors may occur in low-dimensional manifold. Continuity errors mean the nearby neighbourhood data points in the original space can be projected further away in the projected space; they cause projected space to tear. On the other hand,

trustworthiness methods, such as CCA and CDA, depend on the point relations in low-dimensional manifold, rather than those relations in original space, to find low-dimensional manifold. The points coordinates can be updated through a projection process in low-dimensional manifold in a flexible way without constraints to preserve the structure of original space. CCA and CDA might face false neighbourhoods in their low-dimensional manifold [18].
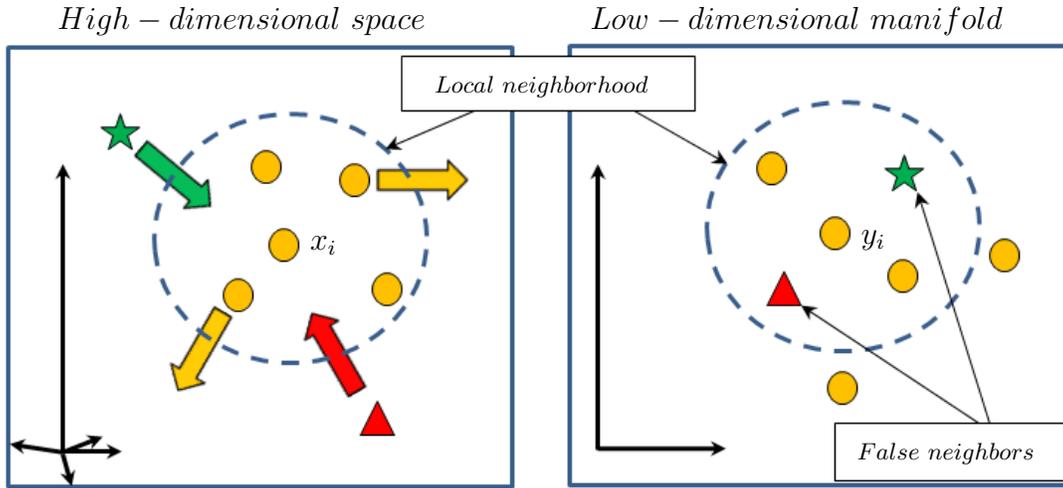

Figure 8: *DR might cause the points which are outside local neighborhood in high dimension space to be inside local neighborhood in low dimension manifold. These points are called false neighborhood.*

In false neighbourhoods, the farther away data points in original space can be projected nearby in the projected space; they cause the projected space to be overlapping, as in Fig.8. Some techniques, as in [19] and [20], are used to overcome the false neighbourhood problems by focusing on the projected space distance to preserve the original distance. The researchers in [19] fixed the false neighbourhoods by sending some of them away to improve the trustworthiness of projected space. Although some regions of their projected space are improved by this method, the final visualization might be tearing because the discarded data points left holes in their location when sending them away. In visualization, the false neighbourhoods are more dangerous than the continuity errors because of the points that take the incorrect colours are unrelated to the original information.

4.　Quality of Visualization

When measuring the quality of visualization for a given data sets, it is important to know which DR method is suitable for the task at hand. Furthermore, the user cannot compare the quality of a given visualization with the original data by visual inspection due to its high-dimensionality. Thus, the formal measurements should evaluate the amount of the preserving neighbourhood colour distances in the visualization with their corresponding distance in original data. Correlation (γ), local continuity (LC) and point-wise quality ($\gamma(i)$) are the good metrics can be used in this matter. If we suppose X is a vector of all points of the data sets in original space and Y is a vector of all the corresponding points in projected space. A and B are the vectors of all pairwise distance of X and Y, respectively, then:

Correlation function (γ): this metric computes the linear correlation between original input distances and colour distances in visualization [21]. The value of correlation is equal to 1 when all distances are perfect preserved, where positive slope between two vectors with perfect linear. In the other hand, the value equal to -1 if the two vectors have prefect linear relationship with negative slope. The correlation metric is defined as the follow:

$$\gamma = \frac{A^T \ B/|A| - \overline{A} \ \overline{B}}{\sigma_A \ \sigma_B} \tag{20}$$

where $|A|$ is the number of components in A, and $\overline{A}$ and σ_A are the mean and standard deviation of A, respectively.

Local Continuity (LC) computes the degree of similarity between two corresponding nearest neighbours sets in

projected and original spaces [22]. The average of all cases represents the efficiency measurement of a projected space. Formally, let k nearest neighbours set in original space to data point i is $\mathcal{N}_k^X(i) = \{j_1, j_2, ..., j_k\}$, and the k nearest neighbours set to i in projected space is $\mathcal{N}_k^Y(i) = \{l_1, l_2, ..., l_k\}$. The measurement of overlapping between two sets is evaluated by

$$N_k(i) = |\mathcal{N}_k^X(i) \cap \mathcal{N}_k^Y(i)| \tag{21}$$

$N_k(i)$ is normalized to the [0,1] interval in order to compute the faithfulness measure of i:

$$Faithfulness_k(i) = \frac{1}{k} N_k(i) \tag{22}$$

In this case the faithfulness value of projected space will be :

$$Faithfulness_k = \frac{1}{N} \sum_{i=1}^{N} Faithfulness_k(i) \tag{23}$$

where N is total number of data points in data sets.

Point-wise Quality Metric ($\gamma(i)$): it computes the correlation for each point through the following equation:

$$\gamma(i) = \frac{A_i^T \ A_i/|A_i| - \overline{A}_i \ \overline{B}_i}{\sigma_{A_i} \ \sigma_{B_i}} \tag{24}$$

where A_i is a vector of all pairwise distance of the data point i to the all other points in original space and B_i is a vector of its corresponding pairwise colour distances in the visualization [14]. The result of this point-wise metric, in Equation 24, is a faithful image, which represents the degree of matching the colour distances in visualization with their corresponding distances in the bands of a remote sensing imagery data sets. It identifies the weak and strong faithful pixels, where the very dark pixels have the lowest faithful values, and the white pixels have the highest faithful values. The number of faithful colours will be very high in faithful images when the colour distances among pixels in visualization and their corresponding distance in original spaces are preserved. On the other hand, the faithful image has a lot of points that have low faithful values to indicate when the visualization has false colours, because their colour distances among pixels are not preserved with their corresponding pixels in the original space.

5.   Conclusion

The objective of using DR is to transfer the high-dimensional data sets to the low-dimensional space so that it is maintaining the essential original information. The simplest approach is to use linear methods, but the complexity of the recent data sets make these methods useless. Thus, the nonlinear versions of those linear methods were introduced to overcome their limitations. Nonlinear transformations can be used to project a high-dimensional data sets in a low-dimensional space. Although this method succeeded in solving some problems, some nonlinear DR methods perform poorly on complex data sets, such as LLE. Nonlinear DR methods, such as Isomap and CDA, which are used in geodesic distance have proven their ability to deal with different complex data sets. The consuming time and storage when using geodesic distance reduces their efficiency. The benefit of visualizing by *TSPE* is that it is able to recognize the features by preserving their point distances between the projected space and the original data sets. *TSPE* can overcome many of the problems introduced by false neighbourhood by deriving higher quality point relationships in its low-dimensional representation. The *TSPE* is better than many DR methods because the *TSPE* prevents false neighbourhood errors to occur in the results.

References

[1] J. Zhang, Visualization For Information Retrieval, Springer-Verlag Berlin Heidelberg, 2008.

[2] I. Borg, P. Groenen, Modern Multidimensional Scaling: Theory and Applications., Springer Verlag, 2005.

[3] S. Nishisato, Multidimensional Nonlinear Descriptive Analysis, Boca Raton, FL: Chapman & Hall, 2006.

[4] J. A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer, 2007.

[5] L. Yang, Distance-preserving dimensionality reduction., Wiley Interdisc. Rew.: Data Mining And Knowledge Discovery 1 (2011) 369–380.

[6] D. K. Agrafiotis, Stochastic proximity embedding, Computational Chemistry 24 (2003) 1215–1221.

[7] I. T. Jolliffe, Principal Component Analysis, Springer Verlag, New York, Inc., 2002.

[8] J. B. Tenenbaum, V. de Silva, J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (2000) 2319–2323.

[9] J. Clark, D. A. Holton, A First Look At Graph Theory, World Scientific Publishing Co. Pte. Ltd., 2005.

[10] O. Samko, A. Marshall, P. Rosin, Selection of the optimal parameter value for the isomap algorithm, Pattern Recognition Letters 27 (2006) 968–979.

[11] X. Zeng, S. Luo, Generalized locally linear embedding based on local reconstruction similarity, in: Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008.

[12] J. A. Lee, A. Lendasse, M. Verleysen, Nonlinear projection with curvilinear distances: Isomap versus curvilinear distance analysis, Neurocomputing 57 (2004) 49–76.

[13] J. X. Li, Visualization of high-dimensional data with relational perspective map, Information Visualization 3 (2004) 49–59.

[14] S. A. Najim, I. S. Lim, Trustworthy dimension reduction for visualization different data sets, Information Sciences 278 (2014) 206–220.

[15] S. Kaski, J. Peltonen, Dimensionality reduction for data visualization, IEEE Signal Processing Magazine 28 (2011) 100–104.

[16] A. Telea, Data Visualization Principles and Practice, A K peters, Ltd, 2008.

[17] T. Schreck, T. von Landesberger, S. Bremma, Techniques for precision-based visual analysis of projected data., in: IS & T/SPIE Conference on Visualization and Data Analysis (VDA 2010) . San Jose, California., 2010.

[18] S. Lespinats, M. Aupetit, CheckViz: Sanity check and topological clues for linear and non-linear mappings, Computer Graphics Forum 30 (2011) 113–125.

[19] S. Kaski, J. Nikkil, M. Oja, J. Venna, P. Trnen, E. Castrn, Trustworthiness and metrics in visualizing similarity of gene expression, BMC Bioinformatics 4 (2003) 1–13.

[20] J. Venna, S. Kaski, Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity, in: WSOM Conference 05. Paris, France, 2005.

[21] M. Mignotte, A bicriteria optimization approach based dimensionality reduction model for the color display of hyperspectral images, IEEE Transactions on Geoscience and Remote Sensing 50 (2012) 501–513.

[22] L. Chen, A. Buja, Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis, Journal of the American Statistical Association 104 (2009) 209–219.