

# Assessment of Dimensionality Reduction Based on Communication Channel Model; Application to Immersive Information Visualization

Mohammadreza Babae<sup>\*</sup>, Mihai Datcu<sup>†</sup> and Gerhard Rigoll<sup>\*</sup>

<sup>\*</sup>Institute for Human-Machine Communication,  
Technische Universität München & Munich Aerospace Faculty, Munich, Germany  
Email: {reza.babae,rigoll}@tum.de

<sup>†</sup>Munich Aerospace Faculty, German Aerospace Center (DLR), Wessling, Germany  
Email: mihai.datcu@dlr.de

**Abstract**—We are dealing with large-scale high-dimensional image data sets requiring new approaches for data mining where visualization plays the main role. Dimension reduction (DR) techniques are widely used to visualize high-dimensional data. However, the information loss due to reducing the number of dimensions is the drawback of DRs. In this paper, we introduce a novel metric to assess the quality of DRs in terms of preserving the structure of data. We model the dimensionality reduction process as a communication channel model transferring data points from a high-dimensional space (input) to a lower one (output). In this model, a co-ranking matrix measures the degree of similarity between the input and the output. Mutual information (MI) and entropy defined over the co-ranking matrix measure the quality of the applied DR technique. We validate our method by reducing the dimension of SIFT and Weber descriptors extracted from Earth Observation (EO) optical images. In our experiments, Laplacian Eigenmaps (LE) and Stochastic Neighbor Embedding (SNE) act as DR techniques. The experimental results demonstrate that the DR technique with the largest MI and entropy preserves the structure of data better than the others.

**Keywords**—Dimensionality Reduction; Immersive information Visualization; Communication channel; Quality Assessment

## I. INTRODUCTION

The volume of multimedia data in different applications is increasing exponentially since the last decade. For instance, the amount of collected EO images is increasing in the order of hundreds of terabytes a day. To handle this amount of data, modern data mining techniques represent image features by high-dimensional feature descriptors (e.g., SIFT [1] and Weber [2], etc). Although high-dimensional features improve the performance of data mining algorithms, they make knowledge discovery more complex.

Information visualization gains an increasing interest due to the ability to provide humans with a visual representation of data, especially for presenting the visual insight into unstructured or weak-structured data. However, many visualization techniques have limitations in dealing with the large-scale high-dimensional data. For example, parallel coordinates are well known for visualization of multi-dimensional data, but they suffer from scalability [3]. Furthermore, the dimension of the display screen is limited which limits the dimensionality of represented data. Considering the aforementioned problems, immersive information visualization system can be seen as an

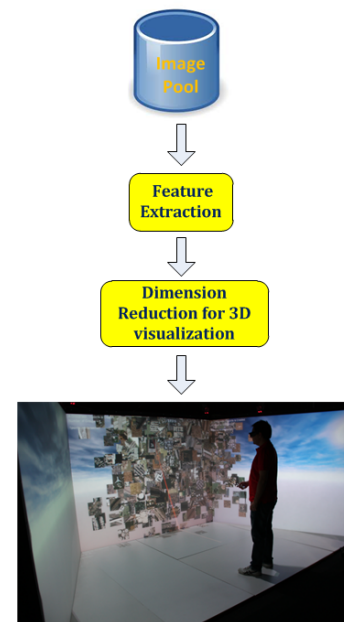


Fig. 1. An immersive visualization system provides the user with a visual representation of data. Here, high-dimensional features are extracted from a database of Earth Observation images and are fed into a dimension reduction technique to be visualized in an immersive 3D virtual environment.

alternative. See Fig. 1. In this system, the high-dimensional data is projected into three-dimensional space by DR techniques. This allows the data to be visualized in an immersive virtual 3D environment. The drawback of using DR techniques for visualization is information loss due to reducing the dimensionality of data. In fact, the information loss depends on the used DR techniques and selecting their parameter(s). Therefore, it is important to assess DR algorithms in order to choose the most suitable one with proper parameter(s).

In this paper we present a new metric based on a communication channel model to assess the quality of dimensionality reduction for visualization. Indeed, we consider the dimensionality reduction as a communication channel in which high-dimensional data points (the input of the channel) are transferred into a low-dimensional space (the output of the channel). We claim that the mutual information and the entropy

of the probability distribution defined over the co-ranking matrix measures the quality of the used DR technique in terms of preserving the structure of data during dimension reduction process.

The rest of the paper is organized as follows. In Section II, we present an overview of the used DR techniques followed by reviewing several state-of-the-art quality measures. We describe our proposed approach followed by the concept of immersive visualization in Section III. Section IV explains experimental set-up and results, and finally, Section V presents the conclusion and future works.

## II. RELATED WORK

### A. Dimensionality Reduction

Since the last two decades numerous linear and nonlinear DR techniques have been proposed in different research areas. While linear approaches assume the data comes from a linear  $d$ -dimensional subspace of a high-dimensional feature space, nonlinear approaches consider the data as a  $d$ -dimensional manifold embedded in a high-dimensional space. Perhaps, the most famous linear algorithm is Principal Component Analysis (PCA) projecting data into  $d$  eigenvectors corresponding to  $d$  largest eigenvalues of the covariance matrix of the data.

Nonlinear methods work based on the minimization of an objective function in which some constraints are involved. Among the nonlinear methods, Locally Linear Embedding (LLE) [4] aims to preserve the topology of data during dimension reduction. It assumes that the data belongs to a low-dimensional smooth and nonlinear manifold embedded in a high-dimensional space. Then the data points are mapped to a lower dimensional space in such a way as the computed linear combination of the data points and their neighbors is preserved.

Laplacian Eigenmaps (LE) [5] is another nonlinear technique in the domain of spectral decomposition methods. It accomplishes the task by building a neighborhood graph from the given data whose nodes represent data points and edges depict the proximity of neighboring points. This graph approximates the low-dimensional manifold embedded in a high-dimensional space. The eigenfunctions of the Laplace Beltrami operator on the manifold serve as the embedding dimensions.

Stochastic Neighbor Embedding (SNE) [6] is a probabilistic approach aiming to preserve the neighborhood of data points by working on probability instead of distances between points. More precisely, the neighborhood relation of data points in both high and low-dimensional spaces is represented by a probability matrix, in which the closer neighboring points have larger probability values than the farther ones. The sum of the Kullback-Leibler divergences over the probability matrices is used as the cost function in the minimization procedure.

The output of dimensionality reduction varies significantly with the used technique and selected parameter(s). For instance, two different techniques with the same input and parameter exhibit completely different results. Fig. 2 shows the visualizations of an optical data set, where the dimensionality of the extracted high-dimensional Weber descriptors (WLD) is reduced by two different DR techniques, namely, LE and SNE.

### B. Quality Assessment

Various dimension reduction techniques have been developed, which exhibit different results based on the input data and tuning parameter(s). Therefore, it is necessary to evaluate their quality in order to choose a proper one with the right parameter(s) for data mining systems.

Since the majority of DR techniques focus on preserving the local neighborhood distances between data points, state-of-the-art approaches try to improve on the succession of DR techniques in preserving the distances. These approaches can be categorized to four categories. The first group evaluates the performance of DR by the assessment of the value of the cost function after convergence [5], [7]. Clearly, these approaches are useful to compare the results of a specific technique with different set of parameter(s).

The second group focuses on the reconstruction error [8]. However, since the reverse transformation does not exist for all techniques, it is hard to employ these approaches for all DR techniques.

The third group judges DR techniques based on the accuracy of classification applied on labeled data [9]. The main drawback of this group is the need for labeled data which is not available in most cases.

Finally, the last group comprises approaches concentrating on preserving the structure of data. The current criteria for the assessment of the preservation of data structure are the local continuity meta-criterion (LCMC) [10], [11], the trustworthiness and continuity measures (T&C) [12], and the mean relative rank error (MRRE) [13], [14]. All these criteria analyze the neighborhoods before and after the dimension reduction. A recent work has put all these criteria to a single framework to compare them [14]. The advantage of this framework is its ability to propose new criteria for the assessment of DR techniques.

## III. PROPOSED APPROACH

### A. Communication Channel Model

Modeling the information transmission in a processing pipeline as a communication channel has been widely used in different research areas [2], [15]. In our paper, dimensionality reduction is modeled as a communication channel in which data points from a high-dimensional space are transferred into a low-dimensional space. Therefore, measuring the quality of this channel reflects the quality of the used dimension reduction technique. Evidently, knowing the fact that recent approaches in DR attempt to preserve the structure of data during dimension reduction, we encode the structure of data in a matrix, the so-called ranking matrix [14].

1) *Ranking Matrices*: if we define a data point in a high-dimensional space as  $X_i = [x_{i1}, \dots, x_{iN}]$  and its correspondence in a low-dimensional space as  $Y_i = [y_{i1}, \dots, y_{id}]$ , where  $d \ll N$ , the ranking matrices of the data points before and after dimensionality reduction are  $A_{ij}$  and  $B_{ij}$ , respectively, where

$$A_{ij} = |\{\kappa : \alpha_{i\kappa} < \alpha_{ij} | (\alpha_{i\kappa} = \alpha_{ij} \ \& \ \kappa < j)\}| \quad (1)$$

$$B_{ij} = |\{\kappa : \beta_{i\kappa} < \beta_{ij} | (\beta_{i\kappa} = \beta_{ij} \ \& \ \kappa < j)\}| \quad (2)$$

where  $|\cdot|$  gives the set cardinality. The  $ij$ th element of this matrix shows how many data points are closer to the point  $i$  than the point  $j$ . Due to the change of distances between data

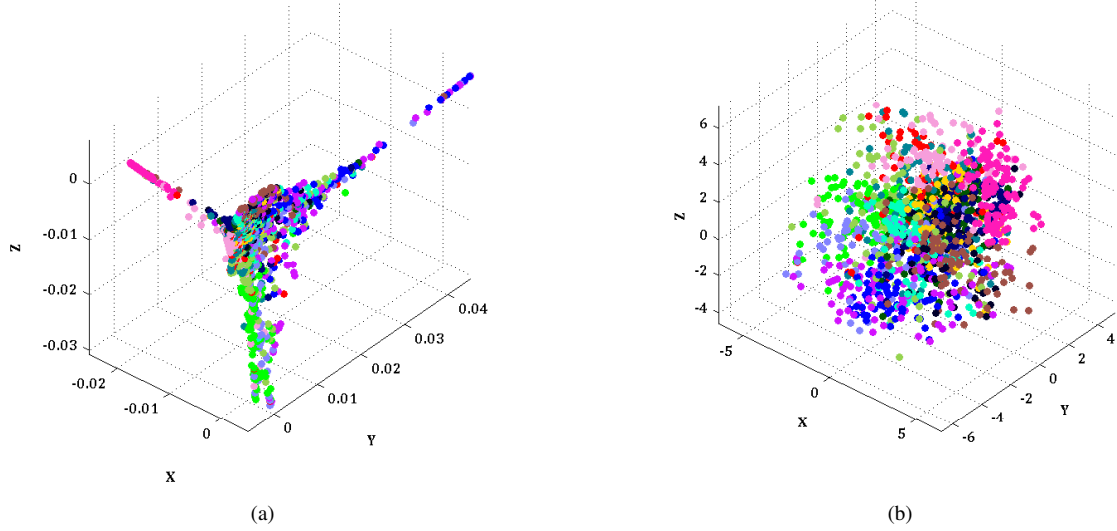


Fig. 2. Two dimension reduction techniques exhibit different results with the same input and parameter. High-dimensional Weber Local descriptors (WLD) extracted from Corel data set are fed into LE and SNE algorithms to reduce the dimensions for visualization. a) Result from LE; b) Result from SNE.

points during the dimension reduction, the ranking matrix of high-dimensional data points (input ranking matrix) changes to the ranking matrix of low-dimensional points (output ranking matrix). Since the ranking matrices can be imagined as 2D images, image similarity measures can be employed to quantify the similarity degree of the input and output ranking matrices. One of the widely used image similarity measures is the mutual information of the probability distribution defined over the joint histogram of ranking matrices.

2) *Co-ranking Matrix*: The similarity of ranking matrices is represented by the joint histogram of input and output ranking matrices, namely, the co-ranking matrix [16], which is defined as

$$M = [m_{kl}]_{1 \leq k, l \leq N-1} \quad (3)$$

for  $N$  data points and

$$m_{kl} = |\{(i, j) : (A_{ij} = k) \& (B_{ij} = l)\}|. \quad (4)$$

3) *Mutual Information*: Mutual information (MI) and entropy are common similarity measures that can be applied to reflect the similarity degree of ranking matrices. But first a joint probability distribution, namely  $P(i, j)$ , should be defined over the co-ranking matrix by

$$P(i, j) = \frac{1}{N-1} M \quad (5)$$

Therefore, the entropy is defined by

$$H = - \sum_i \sum_j P(i, j) \log P(i, j) \quad (6)$$

and the mutual information is:

$$MI = \sum_i \sum_j P(i, j) \log \frac{P(i, j)}{P(i)P(j)} \quad (7)$$

Obviously, when a DR technique completely preserves the structure of data points, both ranking matrices are similar and

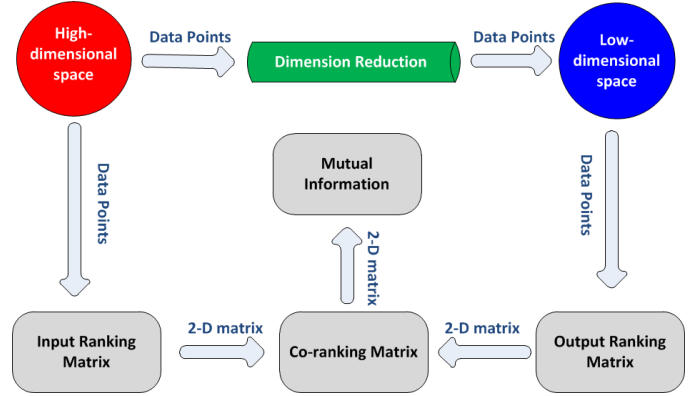


Fig. 3. The workflow of the proposed approach. While data points are transferred from a high-dimensional space into low-dimensional one, the ranking matrices are built from the data points. These matrices are merged together to build up the co-ranking matrix that is used to define a joint probability distribution. Mutual information computed from this probability distribution is used to assess the quality of dimension reduction (here, communication channel).

aligned together. Consequently, the co-ranking matrix would be a diagonal matrix with  $N - 1$  on diagonal values. In this case, the mutual information has its maximum value.

### B. Immersive Visualization and Interaction

Besides numerical evaluation of DR techniques, we propose an immersive 3D virtual environment for visualization of data points. This environment, the so-called Cave Automatic Virtual Environment (CAVE) is based on Virtual Reality technology and comprises four room-sized walls aligned to form a cube to display the low-dimensional features. This configuration allows users to have a 180 degree horizontal view. The virtual scene is projected onto the walls using two projectors per wall in order to have stereoscopic scenarios. Additionally, a real-time tracking system including six infrared cameras

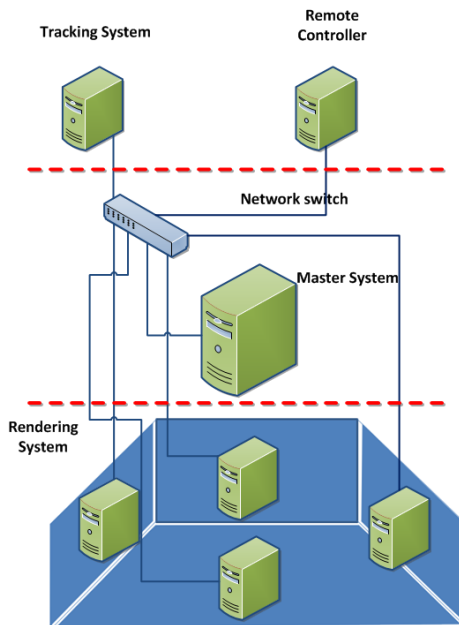


Fig. 4. The physical diagram of immersive visualization. The visualization system is composed of three layers with different responsibilities. The first layer comprises motion capture (tracking) system and control capturing. A master PC in the middle layer for the synchronization, and finally four systems for rendering for each wall of the CAVE. All systems are connected together via an ethernet network.

mounted on top of the walls computes the pose (position and orientation) of marked objects (e.g., Wii controller and shuttle glasses) inside the cube.

For rendering and visualizing the data, the CAVE utilizes a three-layer cluster of PCs. The first layer captures user motions and navigation signals and sends them to the middle layer. Motion capturing is performed by an optical tracking system and the navigation signals are generated by a Wii controller. The middle layer comprises a master PC which is responsible for generating the virtual scene based on the incoming signals from the first layer. Once the scene is ready, it sends rendering signals to the third layer. Rendering and displaying the scene on the walls is carried out by four PCs (one for each wall). The schematic of the organization of the CAVE is depicted in Fig. 4.

#### IV. EXPERIMENTAL RESULTS

To validate our proposed approach we made an experimental setup and then evaluated and visualized the output of the used dimension reduction techniques. In the following we explain our experiments in more details.

##### A. Set-up

Our experimental set-up comprises data sets, features extraction, and dimensionality reduction.

1) *Datasets*: We used two data sets, the first one is UC Merced-Land-Use data set comprising 2100 images categorized in 21 groups. Each group contains 100 image patches of the size  $256 \times 256$  pixels from aerial photography. These images are collected such that they represent rich variation of

scene patterns. Therefore, they are homogeneous in terms of color distribution and texture. Fig. 5 shows some samples of this data set.

The second data set is Corel image data set. This data set contains 1500 images in 15 different groups, where each group contains 100 images. Fig. 6 depicts some sample images from this data set.



Fig. 5. UC Merced-Land-Use is a manually labeled dataset containing 21 classes of land-use scenes: Agricultural, Airplane, Baseball Diamond, Beach, Buildings, Chaparral, Dense Residential, Forest, Freeway, Golf Course, Harbor, Intersection, Medium Density Residential, Mobile Home Park, Overpass, Parking Lots, River, Runway, Sparse Residential, Storage Tanks, Tennis Court. Each image represents one sample of each group.



Fig. 6. Corel images is a manually labeled dataset containing 15 classes of different scenes: Africa, Beach, Bus, Card, Dino, Elephant, Flower, Food, Grote, Horse, Mountain, Portrait, Rome, Sunset, Tiger. Each image represents one sample of each group.

2) *Feature Extraction*: Three different features; namely, color-histogram [17], spectral-SIFT [1], and spectral-WLD [2], [18] are extracted from these images. The extracted feature descriptors are represented by the bag-of-words model; where each image is described by a vector of 200 visual words.

3) *Dimensionality Reduction*: We apply three different dimensionality reduction techniques on our high-dimensional data to reduce the dimensionality to 3D for visualization. These techniques are: Laplacian Eigenmaps (LE) [5], Stochastic Neighbor Embedding (SNE) [6], and Locally Linear Embedding [4].

##### B. Evaluation and Visualization

Mutual information and entropy of the co-ranking matrix are computed for 9 different combinations of features-DR (namely, color-LE, color-SNE, color-LE, sift-LE, sift-SNE, sift-LLE, weber-LE, weber-SNE, weber-LLE) for both Merced and Corel data sets. The computed mutual information and entropy from co-ranking matrices of these combinations

are depicted in Fig. 7.a and Fig. 7.b for Merced and Corel data sets, respectively. The Fig. 7.c and Fig. 7.d show the 3D plot of the used SNE dimension reduction applied to extracted Weber features from Merced and Corel data sets, respectively. By looking at the corresponding mutual informations (values for number 8 in Fig. 7.a and Fig. 7.b) we conclude that larger mutual information exhibits better clustering performance. This is mainly due to the direct relation between DR and clustering. Fig. 8 depicts some samples of immersive visualization of extracted features and their corresponding images for Corel data set. We used immersive visualization because it is much more convenient for users to detect patterns in data set. Users can freely walk in the virtual environment and look at the data set with different viewing angles. The evaluation of our visualization system will be discussed in our future work.

## V. CONCLUSION

In this paper we proposed a new criterion for the assessment of dimension reduction techniques as part of immersive information visualization. We suggested to model the dimension reduction as a communication channel that transfers the data points from a high-dimensional space into a low-dimensional one. In order to assess the quality of this channel, we considered the side effect of dimensionality reduction on ranking matrices represented by the co-ranking matrix. Having defined a joint probability distribution over this matrix we were able to measure the similarity degree of the structure of the input data and the output data via mutual information. Unlike local assessment of DR techniques, our proposed approach globally evaluates the performance of DR by measuring its structure preserving property.

## ACKNOWLEDGMENT

The authors would like to thank Munich Aerospace Faculty, German Aerospace Center (DLR) for their support for this research.

## REFERENCES

- [1] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, vol. 2. Ieee, 1999, pp. 1150–1157.
- [2] R. Bahmanyar and M. Datcu, "Measuring the semantic gap based on a communication channel model," in *Image Processing, IEEE International Conference on*, 2013.
- [3] J. Choo, H. Lee, Z. Liu, J. Stasko, and H. Park, "An interactive visual testbed system for dimension reduction and clustering of large-scale high-dimensional data," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2013, pp. 865 402–865 402.
- [4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [5] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [6] G. Hinton and S. Roweis, "Stochastic neighbor embedding," *Advances in neural information processing systems*, vol. 15, pp. 833–840, 2002.
- [7] M. Bernstein, V. De Silva, J. C. Langford, and J. B. Tenenbaum, "Graph approximations to geodesics on embedded manifolds," Technical report, Department of Psychology, Stanford University, Tech. Rep., 2000.
- [8] M. Balasubramanian and E. L. Schwartz, "The isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, pp. 7–7, 2002.
- [9] L. Van der Maaten, E. Postma, and H. Van den Herik, "Dimensionality reduction: A comparative review," *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.
- [10] L. Chen and A. Buja, "Local multidimensional scaling for nonlinear dimension reduction, graph layout and proximity analysis," Ph.D. dissertation, Citeseer, 2006.
- [11] —, "Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis," *Journal of the American Statistical Association*, vol. 104, no. 485, pp. 209–219, 2009.
- [12] J. Venna and S. Kaski, "Local multidimensional scaling," *Neural Networks*, vol. 19, no. 6, pp. 889–899, 2006.
- [13] J. A. Lee and M. Verleysen, *Nonlinear dimensionality reduction*. Springer, 2007.
- [14] —, "Quality assessment of dimensionality reduction: Rank-based criteria," *Neurocomputing*, vol. 72, no. 7, pp. 1431–1443, 2009.
- [15] M. Chen and H. Jaenicke, "An information-theoretic framework for visualization," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 16, no. 6, pp. 1206–1215, 2010.
- [16] J. A. Lee and M. Verleysen, "Rank-based quality assessment of nonlinear dimensionality reduction," in *Proceedings of 16th European Symposium on Artificial Neural Networks*, 2008, pp. 49–54.
- [17] K. E. van de Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 9, pp. 1582–1596, 2010.
- [18] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, and M. Pietikainen, "A robust descriptor based on weber's law," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–7.



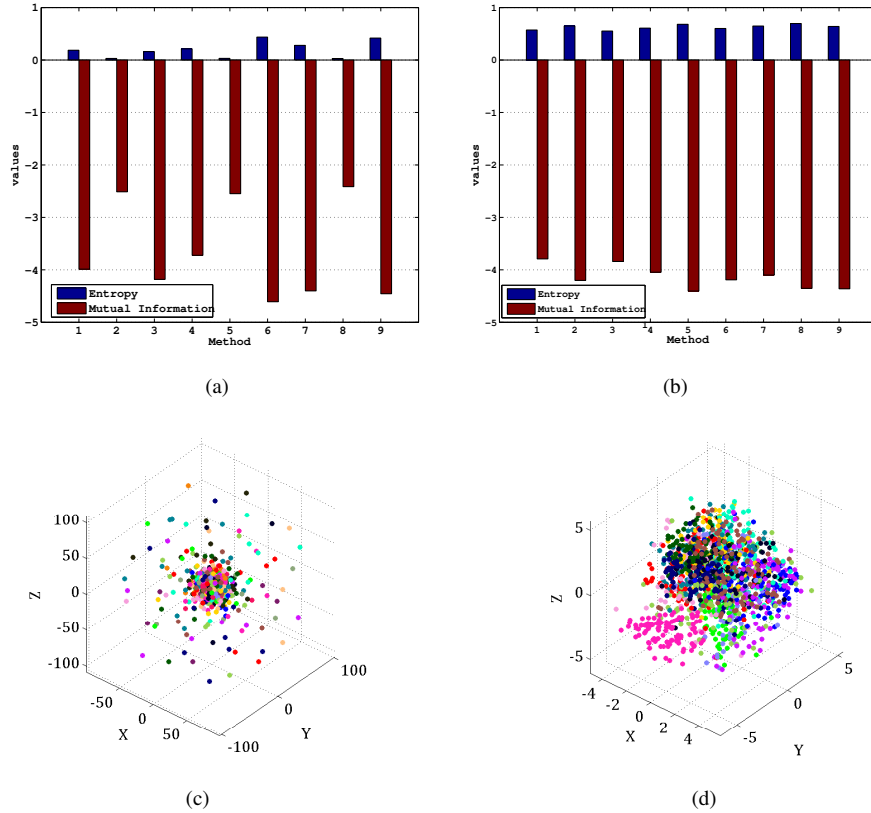


Fig. 7. The quality of dimensionality reduction applied to Merced and Corel datasets represented by mutual information and entropy of co-ranking matrix. A combination of three different features (color-histogram, spectral-SIFT, and spectral-Weber) and three different DR techniques (LE, SNE, LLE) gives 9 feature-DR items which are: 1)color-LE, 2)color-SNE, 3)color-LE, 4)sift-LE, 5)sift-SNE, 6)sift-LLE, 7)weber-LE, 8)weber-SNE, 9)weber-LLE; a) Results on Merced dataset; b) Results on Corel dataset; c) plotted result of method 8 from merced dataset; d) plotted result of method 8 on corel dataset.

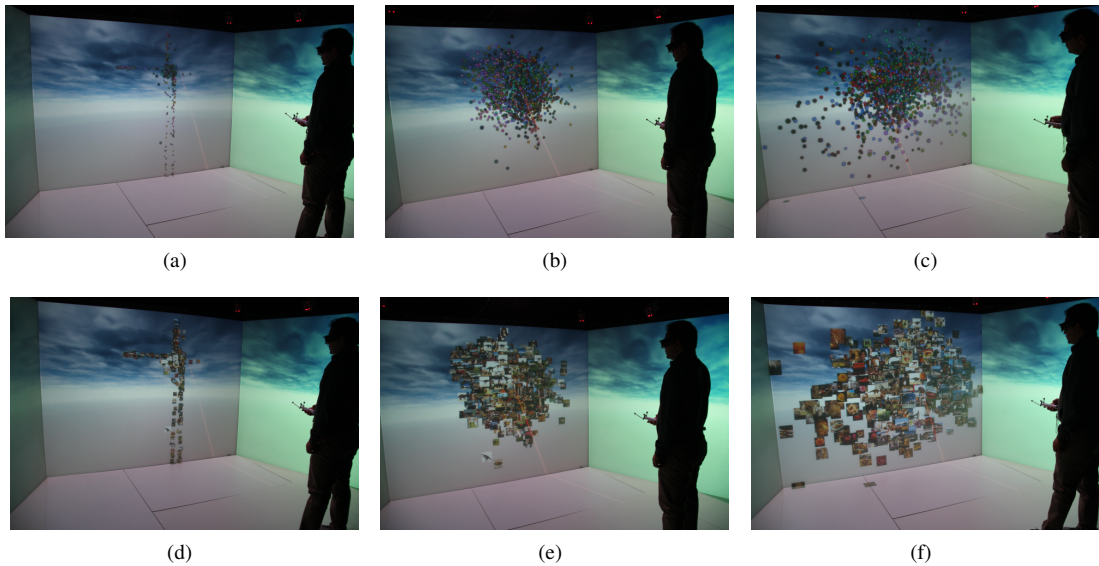


Fig. 8. Some sample images of immersive visualization dimensionality reduction applied on Corel data set. First row depicts the results with images and second row depicts the results as features. (a,d) depict the results of color-LE. (b,e) depict the result of color-SNE. (c,f) depict the result of sift-SNE