2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Geometry Constrained Sparse Embedding for Multi-dimensional Transfer Function Design in Direct Volume Rendering

Zhenzhou Shao[1], Yong Guan[2], Hongsheng He[1] and Jindong Tan[1]

***Abstract*— Direct volume rendering (DVR) is commonly employed for the medical visualization. Multi-dimensional transfer functions are used in DVR to emphasize the region of interest in details. However, it is impractical to interact directly with the functions in more than three dimension. This paper proposes a novel framework called geometry constrained sparse embedding (GCSE) for dimensionality reduction (DR). GCSE allows the conventional DR methods to be applied to a dictionary with much smaller atoms instead. The mapping derived from the dictionary feeds to the original features to obtain the ones in the reduced dimension. To obtain a good dictionary, the intrinsic structure of features is encoded in the sparse embedding based on a geometry distance. In addition, stochastic gradient descent algorithm is employed to speed up the dictionary learning. Various experiments have been conducted using both synthetic and real CT data sets. Compared with conventional methods, GCSE not only produces the comparable results, but also performs well with the capability to handle the large data set more powerfully. The rendering results using the real CT data has demonstrated the effectiveness of GCSE.**

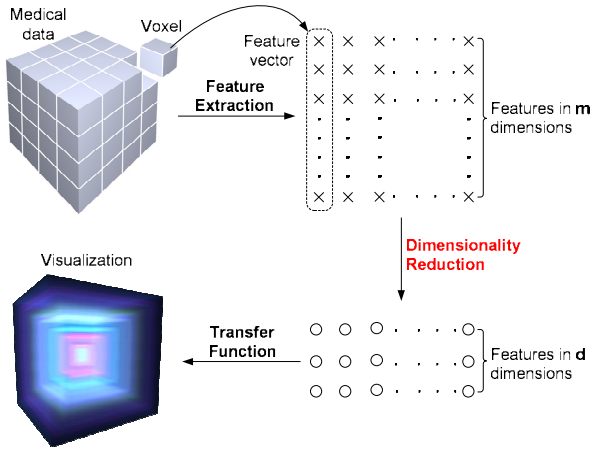


Fig. 1: Dimensionality reduction for transfer function design.

I. Introduction

Efficient interpretation and perception of 3-D medical data is crucial to surgeons during the clinical intervention [1]. Commonly the direct volume rendering (DVR) technique is employed for visualization by projecting the volume data in a projected image without the data misclassification in the segmentation stage [2]. DVR is also used in other procedures, such as surgical training, diagnosis and surgical planning.

In order to define the meaningful visualization and emphasize the regions of interest in the volume data, transfer functions (TFs) are used for mapping from voxels with the scalar information to color and optical properties [3]. Compared with the simple TF using the scalar value only, multi-dimensional TFs are commonly designed based on several features, such as gradient magnitude, curvature and statistical measures, to enable more accurate perception for surgeons.

As shown in Fig. 1, multiple features in TF domain are extracted, and stored in a vector every voxel. However, as the dimensionality of TF domain increases, the direct interaction with the transfer function becomes impractical. In this case, we have to find a low-dimensional representation of high-dimensional features to enable the effective interaction. One solution to this problem is dimensionality reduction (DR) which is the transformation from the high-dimensional data to a representation in the reduced dimension without significant information loss [4]. Another advantage of dimensionality reduction is that the redundant information can be filtered out simultaneously.

Conventional approaches of dimensionality reduction are primarily divided into two categories: linear and non-linear methods. Linear dimensionality reduction is characterized by the linear mapping, e.g. principal component analysis (PCA) [5], and the nonlinear methods, e.g. Isomap [6] and locally linear embedding (LLE) [7], perform better by taking into account the nonlinearity of the original data. However, both methods incur the problem of heavy computational load on the large data set. For example, assume the dimension of TF domain is 5, and a CT volume with the size of $256 \times 256 \times 128$ is given, the input to dimensionality reduction methods is in $\mathbb{R}^{5\times 8388608}$. Nevertheless, the fast implementation of transfer function design is required in both pre- (diagnosis and surgical planning) and intra-operation.

Recently, the rapid development in the filed of sparse representation (SR) paves another way to the dimensionality reduction. This method is achieved based on the assumption that the signals are compressible, and have a sparse representation in a basis set (also a.k.a dictionary). It has been proven practical in many applications [8]–[10]. In [11], Nguyen *et al.* proposed a sparse embedding (SE) framework by combing dimensionality reduction and sparse learning together. It enables the sparse representation (sparse coding and dictionary learning) in the reduced dimension with lower computational load. However, current SR based methods are

[1]Zhenzhou Shao, Hongsheng He and Jindong Tan are with the Department of Mechanical, Aerospace, and Biomedical Engineering, The University of Tennessee, Knoxville, TN, 37996, USA {zshao,hhe,tan}@utk.edu

[2]Yong Guan is with the Beijing Key Laboratory of Electronic System Reliability Technology, Capital Normal University, Beijing, 100048, China guanyong@mail.cnu.edu.cn

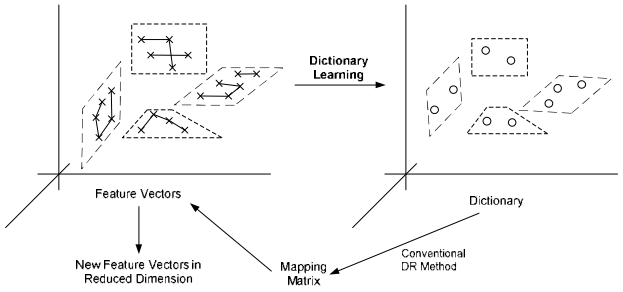


Fig. 2: Illustration of geometry constrained sparse embedding.

still time-consuming as the mapping matrix is optimized together with SR, especially for the medical data with the large size.

The objective of this paper is to propose a dimensionality reduction method with the capability to process the large data set efficiently for the transfer function design in the direct volume rendering. In this paper, a novel framework of Geometry Constrained Sparse Embedding (GCSE) is proposed. To improve the efficiency of dimensionality reduction, the mapping matrix is separated from the framework of SR in Nguyen's work, and it is obtained by applying the conventional method (i.e. PCA, Isomap and LLE) to the dictionary under the assumption that if the features vectors can be described well over an overcomplete dictionary, the mapping derived from the dictionary can embed the entire features in the target dimension. Therefore, the problem is reduced to the optimization of dictionary. To this end, the intrinsically geometrical structure based on a geometry distance are taken into account. The geometry distance is to measure the similarity and find the geometry relationship amongst data, while Euclidean distance just computes the pairwise metric between data. Moreover, to achieve faster implementation, the sparse representation in the proposed method is implemented based on the stochastic gradient descent [12].

II. Geometry Constrained Sparse Embedding

As aforementioned, the transfer function design in DVR is reduced to the dictionary learning. To obtain a better dictionary, the intrinsically geometric structure of feature vectors are taken into account in GCSE in this paper.

As shown in Fig. 2, for each datum in a high dimensional data set, a group of neighbors are selected according to a geometry distance of the datum to the hyperplane in the high dimensional manifold, and a neighbor graph is constructed to encode the geometrical information into GCSE. The learnt dictionary can better describe the features with the similar pattern, so that such dictionary can act as a surrogate for the entire features. The framework of GCSE is summarized in Algorithm (1). Steps 2 and 3 are to be discussed in details in following sections.

Algorithm 1 Overview of GCSE.

Input: CT data**,** number of neighbors K, size of dictionary k**,** initial dictionary $D_0 \in \mathbb{R}^{m\times k}$**.**

1: Extract feature vectors $X \in \mathbb{R}^{m\times n}$ from the medical data.
2: Construct a neighbor graph using the geometry distance.
3: Learn the dictionary D^* based on the objective function encoding the geometrical structure.
4: Apply the conventional DR algorithm to D^* to find the mapping $P \in \mathbb{R}^{d\times m}$.

Output: $Y = PX \in \mathbb{R}^{d\times n}$**.**

III. Neighbor Selection Using Geometry Distance

A neighbor graph is constructed to describe the locally geometrical relationship of the feature vectors. As a measure, Euclidean has shown good performance in the neighbor selection. Nevertheless, Euclidean distance is limited as only the pairwise distance to the target data is considered. It is desired that the local relationships in the cluster are also taken into account. The fundamental problem in neighbor graph construction is how to define the distance in selecting neighbors for each feature vector. In this section, a geometry distance is computed to measure the geometrical similarity of the features. Defining this geometry distance as the neighbor criterion, we can explore the geometry relation of features.

In the feature space $\mathbb{R}^m$ with m dimensions, a parallelotope is determined with column vectors $X = [x_i]_{i=1}^K \in \mathbb{R}^{m\times K}$ as the sides, the volume of which can be computed [13] by

$$\mathrm{Vol}(X) = \det{}^{\frac{1}{2}}(X^T X) = \det{}^{\frac{1}{2}}(G(X)) \tag{1}$$

following the definition of Gramian [14]. We further illustrate the distance from a sample to linear manifolds and judge whether a set of vectors are linearly independent in the space. The Euclidean distance from x_K to the linear manifold spanned by $X_{[1,(K-1)]} = \{x_i\}_{i=1}^{K-1}$ can be represented [13] with the volume of the manifold

$$h^K_{[1,(K-1)]} = \sqrt{\frac{\mathrm{Vol}^2(X_{[1,K]})}{\mathrm{Vol}^2(X_{[1,(K-1]})}}. \tag{2}$$

Following the properties of block matrices, it can be further obtained that

$$\begin{aligned} & \mathrm{Vol}^2(X_{[1,K]}) \\ = \; & \det([X_{[1,(K-1)]}, x_K]^T [X_{[1,(K-1)]}, x_K]) \\ = \; & \det(x_K^T x_K - (x_K X_{[1,(K-1)]})\times \\ & G^{-1}(X_{[1,(K-1)]})(x_K X_{[1,(K-1)]})^T). \end{aligned} \tag{3}$$

According to (1) and (2), we have the following result,

$$\begin{aligned} & h^K_{[1,(K-1)]} \\ = \; & \det(x_K^T x_K - (x_K X_{[1,(K-1)]}) \\ & G^{-1}(X_{[1,(K-1)]})(x_K X_{[1,(K-1)]})^T) \\ = \; & x_K^T x_K - (x_K^T X_{[1,(K-1)]}) \\ & G^{-1}(X_{[1,(K-1)]})(x_K^T X_{[1,(K-1)]})^T. \end{aligned} \tag{4}$$

However, since $K > m$, $X_{[1,(K-1)]}$ may be redundant, so that the inverse matrix of $G(C_{[1,(K-1)]})$ doesn't exist. In this paper, a simple technique to solve this problem is employing the Tikhonov regularization

$$\begin{aligned} & h^K_{[1,(K-1)]} \\ = \; & x_K^T x_K - (x_K^T X_{[1,(K-1)]})^T \\ & (G(X_{[1,(K-1)]}) + \epsilon I)^{-1} x_K^T X_{[1,(K-1)]} \end{aligned} \tag{5}$$

with $\epsilon > 0$ as the regulation parameter. The introduction of ϵ also avoids the inverse of singular Σ whose eigenvalues are near zero in computational implementation. Using distance measurement in Eq. (5), the neighbors of each data point are selected incrementally until the neighbor number is satisfied.

IV. Dictionary Learning

In this section, the dictionary is learnt by encoding the geometric relation of feature vectors, and the stochastic gradient descent method is employed to speed up the implementation of dictionary learning.

A. Objective Function

Let $D = [d_1, d_2 \cdots d_k] \in \mathbb{R}^{m \times k}$ be the dictionary matrix, and $\alpha = [\alpha_1, \alpha_2 \cdots \alpha_n] \in \mathbb{R}^{k \times n}$ denotes the sparse matrix. Given a data set $X = [x_1, x_2, \cdots x_n] \in \mathbb{R}^{m \times n}$, the objective function of original sparse representation can be defined as

$$\begin{aligned} & \{D, \alpha\} = \arg\min_{\{D,\alpha\}} \tfrac{1}{2} \parallel X - D\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 \\ & s.t. \parallel d_i \parallel^2 \leq 1, \; i = 1, \cdots k, \end{aligned} \tag{6}$$

where λ is the regularization parameter. To prevent arbitrarily large values of D, the constraint $\parallel d_i \parallel^2 \leq 1$ is applied. Base on this basic objective function, an extra item is integrated to preserve the geometrical structure of X. To achieve it, we can construct a neighbor graph G with n vertices according to the geometry distance in Section III. Each vertex represents a feature vector of medical data, and it is associated with a element in the weight matrix W of G. If x_j belongs to the neighbors of x_i, $w_{i,j} = 1$, otherwise, $w_{i,j} = 0$. The difference with the graph constructed by Euclidean distance is that the weight matrix W is an asymmetric matrix.

To map the weighted graph G to the sparse coefficients α, a reasonable criterion [15] is chosen by minimizing the following function:

$$\begin{aligned} & \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \parallel \alpha_i - \alpha_j \parallel_F^2 w_{i,j} \\ = \; & \tfrac{1}{2} (\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_i^T w_{i,j} + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_j \alpha_j^T w_{i,j} \\ & -2 \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i^T \alpha_j w_{i,j}) \\ = \; & \tfrac{1}{2} (\mathbf{tr}(\alpha R \alpha^T) + \mathbf{tr}(\alpha C \alpha^T)) - \mathbf{tr}(\alpha W \alpha^T) \\ = \; & \mathbf{tr}(\alpha L \alpha^T), \end{aligned} \tag{7}$$

where $L = \frac{1}{2}(R + C) - W$ is the Laplacian matrix. The degrees of x_i along the row and column are defined as R and C

$$\begin{aligned} & R = diag(r_1, r_2, \cdots r_N), \; r_i = \textstyle\sum_{j=1}^{N} w_{i,j} \\ & C = diag(c_1, c_2, \cdots c_N), \; c_j = \textstyle\sum_{i=1}^{N} w_{i,j}. \end{aligned} \tag{8}$$

By combining Eq. (6) and Eq. (7), we can obtain following objective function:

$$\begin{aligned} & D^* = \arg\min_{D} \tfrac{1}{2} \parallel X - D\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 + \lambda_G tr(\alpha L \alpha^T) \\ & s.t. \parallel d_i \parallel^2 \leq 1, \; i = 1, \cdots k, \end{aligned} \tag{9}$$

where λ_G is the regularization parameter for the geometric structure.

In this paper, stochastic gradient descent is employed to handle the large amount of features of medical data. The dictionary learning algorithm is summarized in Algorithm 2. The dictionary and coefficients can be optimized alternatively by minimizing over one while keeping the other one fixed [16]. To improve convergence rate further, τ feature vectors are selected randomly each iteration. Then the sparse coefficient $\alpha_{t,i}$ of $x_{t,i}$ over the previous dictionary D_{t-1} is computed by minimizing Eq. (10). A and B are calculated to update the dictionary using Algorithm 3. The details are discussed in Section IV-C.

Algorithm 2 Dictionary Learning.

Input: $X \in \mathbb{R}^{m \times n}$, $\lambda, \lambda_G, \lambda_L \in \mathbb{R}$, initial dictionary $D_0 \in \mathbb{R}^{m \times k}$, batch size τ, number of iterations T.

1: $A_0 \leftarrow 0$, $B_0 \leftarrow 0$.
2: **for** $t = 1$ to T **do**
3: Select τ feature vectors $\{x_{t,i}\}_{i=1}^{\tau}$ randomly.
4: Perform sparse coding using feature-sign search algorithm for each vector $x_{t,i}$.

$$\begin{aligned} \alpha_{t,i} = & \arg\min_{\alpha \in R^{k \times 1}} \tfrac{1}{2} \parallel x_{t,i} - D_{t-1}\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 \\ & + \lambda_G \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \parallel \alpha_i - \alpha_j \parallel_F^2 w_{i,j}. \end{aligned} \tag{10}$$

5: $A_t \leftarrow A_{t-1} + \sum_{i=1}^{\tau} (\alpha_{t,i} \alpha_{t,i}^T)$
6: $B_t \leftarrow B_{t-1} + \sum_{i=1}^{\tau} (x_{t,i} \alpha_{t,i}^T)$.
7: Compute D_t based on block-coordinate descent, refer to Section IV-C for details.
8: **end for**

Output: D^*.

B. Sparse Coding

In this section, the sparse coding problem is solved by fixing the dictionary D. According to Algorithm 2, each sparse coefficient vector $\alpha_{t,i}$ is updated individually, while the rest are fixed. In order to optimize $\alpha_{t,i}$ individually, Eq. (9) should be formulated in a vector form. The combination of residual item and sparse regularizer can be denoted as $\sum_{i=1}^{n} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \sum_{i=1}^{n} \parallel \alpha_i \parallel_1$, where the subscript t from x and α is omitted for simplicity. The Laplacian

regularizer $\mathbf{tr}(\alpha L \alpha^T)$ can be expanded as follows:

$$\begin{aligned}\mathbf{tr}(\alpha L \alpha^T) &= \mathbf{tr}\left(\sum_{i,j=1}^{n} L_{i,j}\alpha_i \alpha_j^T\right) \\ &= \sum_{i,j=1}^{n} L_{i,j}\alpha_i^T \alpha_j.\end{aligned} \tag{11}$$

Thus, the problem Eq. (9) can be rewritten as

$$\begin{aligned}\alpha^* &= \arg\min_{\alpha} \tfrac{1}{2}\sum_{i=1}^{n} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \sum_{i=1}^{n} \parallel \alpha_i \parallel_1 \\ &+\lambda_G \sum_{i,j=1}^{n} L_{i,j}\alpha_i^T \alpha_j.\end{aligned} \tag{12}$$

Because only α_i is updated one time, by fixing other sparse coefficients, we can obtain the objective function for α_i as follows:

$$\begin{aligned}\alpha_i^* &= \arg\min_{\alpha_i} \tfrac{1}{2} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \parallel \alpha_i \parallel_1 \\ &+\lambda_G \left(L_{i,i}\alpha_i^T \alpha_i + 2\alpha_i^T \sum_{j\neq i,}^{n} L_{i,j}\alpha_j\right).\end{aligned} \tag{13}$$

In this paper, the above problem is solved using the feature-sign search algorithm [16] as a unconstrained quadratic optimization problem. It outperforms many existing algorithms, such as grafting [17] and LARS [18].

C. Dictionary Update

In this section, the dictionary update is performed using block-coordinate descent by minimizing residual error in Eq. (9). Each column of dictionary is updated sequentially according to Eq. (15) under the constraint $\parallel d_i \parallel^2 \leq 1$. $Da_j - b_j$ is the gradient of $f(D) = \frac{1}{2} \parallel X - D\alpha \parallel_F^2$ with respect to d_j, and η is the learning rate. As mentioned in Section IV-A, the dictionary learning is implemented based on stochastic gradient descent algorithm. Taking partial differentiation with respect to D, we can obtain

$$\begin{aligned}\frac{\partial f}{\partial D} = -(x_i - D\alpha_i)\alpha_i^T &= D(\alpha\alpha^T) - x_i\alpha_i^T \\ &= DA - B,\end{aligned} \tag{14}$$

where $A = \alpha\alpha^T$ and $B = x_i\alpha_i^T$. Since the previous dictionary D_{t-1} is used in the proposed algorithm to provide a prior knowledge for computing D_t, number of iteration T can be set to a small value. In our experiments, only one iteration is enough for the convergence.

V. Experimental Results and Discussions

In this section, three sets of experiments were performed to evaluate the performance of GCSE on both synthetic and real CT data sets. The GCSE firstly employed the synthetic data to compare with the existing methods, including linear and non-linear approaches. The convergence and the robustness of related parameters were also studied. The second experiment was to test the capability of handling a large data set using a partial CT volume extracted from the golden standard data set [19]. For the effect of volume rendering using different transfer functions, there is no quantitative way to measure that. Thus, we presented the rendered images that successfully emphasize the region of interests to compare the difference for visualization. Human tooth CT data was used in the last test.

Algorithm 3 Dictionary update.

Input: $D = [d_1, d_2, \cdots d_k] \in \mathbb{R}^{m\times k}$, $A = [a_1, a_2, \cdots a_k \in \mathbb{R}^{k\times k}] = \sum_{i=1}^{t} \alpha_i\alpha_i^t$, $B = [b_1, b_2, \cdots b_k \in \mathbb{R}^{m\times k}] = \sum_{i=1}^{t} x_i\alpha_i^t$, number of iteration T.

1: **for** $i = 1$ to T **do**
2: **for** $j = 1$ to k **do**
3: Update the jth column of the dictionary:

$$\begin{aligned}d_j &\leftarrow d_j - \eta(Da_j - b_j) \\ d_j &\leftarrow \tfrac{1}{max(\|d_j\|_F, 1)} d_j.\end{aligned} \tag{15}$$

4: **end for**
5: **end for**

Output: D^*(Updated dictionary).

The specifications of data sets used in the experiments are summarized in Table I. Toroidal Helix data set is used directly as the input of dimensionality reduction methods. For the real CT data sets, corresponding feature vectors are both in 6 dimensions. The first vector is related to the intensity for each voxel. To discriminate the voxels next to the boundaries. The following 4 ones are the gradient information, including gradient magnitude and three components in the gradient vector. The last one is the Hessian measure, which can capture the edge information to deal with complex configurations. The target dimension is set to 2 for all experiments.

A. Comparison with Conventional Methods

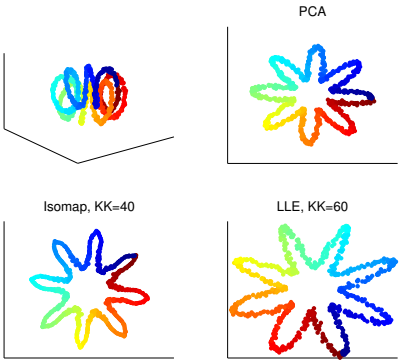
In this experiment, Toroidal Helix data set was used to compared GCSE with conventional methods, including PCA, Isomap and LLE. As shown in Fig. 3a, all conventional methods project the input data into a symmetric and flower-like shape, and preserve the structure and distribution of the helix. The number of neighbors were 40 and 60 for Isomap and LLE, respectively, while PCA does not need any parameters for the dimensionality reduction.

Fig. 3b shows the results of GCSE on Toroidal Helix data set. When the learnt dictionary D was ready, we applied PCA, Isomap and LLE to D, respectively. In this experiment, the size of dictionary was 50, the numbers of neighbors for Isomap and LLE were 3 and 4. We can see that much less neighbors are required to achieve the similar effect compared with the conventional methods. It also preserves the geometric information of the input.

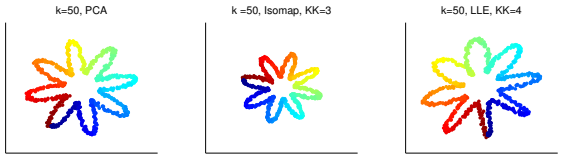
To explain the less number of neighbors in GCSE, Fig. 4 illustrates the distribution of Toroidal Helix data set and

TABLE I: Specifications of synthetic and real CT data sets.

Data source	Size of data set	Size of features	Dim_{target}
Toroidal Helix	3×800	3×800	2
Pig's head [19]	$20 \times 21 \times 20$	6×8400	2
Human tooth	$94 \times 81 \times 155$	6×1180170	2



(a) Conventional methods



(b) GCSE

Fig. 3: Comparison between conventional methods and GCSE.

dictionary used for dimensionality reduction. The dictionary is better able to keep the topological structure with the less atoms, so that the local structure can be presented using less points in the dictionary, while more neighbors are needed to express the same geometric information.

We also studied the convergence and robustness of GCSE based on the stochastic gradient descent algorithm. Following the convergence analysis of stochastic gradient descent in [20], it is obvious that the dictionary can approximate to a stationary point asymptotically. To verify this, we set different sizes of dictionary and numbers of neighbors indicating the geometric structure in the objective function. In the test, the sizes of dictionary were $\{50, 100, 150, 200\}$, and the values for the number of neighbors were $\{5, 10, 15, 20\}$. GCSE is performed with the iteration number of 100, and the batch size in each iteration is also set to 100. The residual errors are shown in Fig. 5. In all cases, the average residual error is 0.29.

B. Comparison with Sparse Embedding

This experiment was implemented to test the capability of large data set handling using Nguyen's SE and GCSE. We tried to do the dimensionality reduction on gold standard data set and human tooth CT. However, SE failed using both data sets due to the memory limit. The same problem occurred when we apply the conventional methods to these data sets. GCSE can deal with this case, the evaluation of GCSE on the large data set is presented in next section.

To compare the performance of GCSE and SE, a partial CT volume extracted from gold standard data was used. The feature vectors are in 6D, and the target dimension is 2. In SE, the iteration number was set to 5, and a

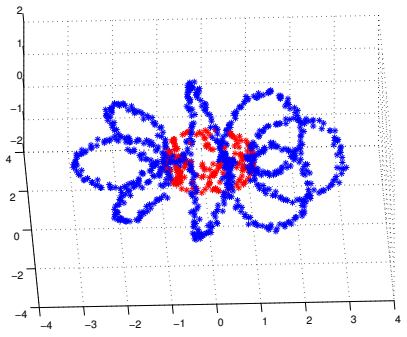
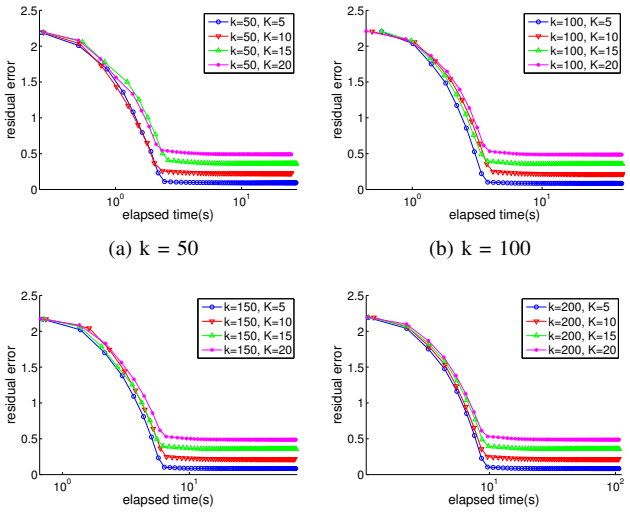


Fig. 4: Topologies of input data and dictionary in 3D. The blue stars are input data, and the learnt dictionary is indicated by the red stars.

polynomial kernel of degree was 4. For GCSE, the size of target dictionary was set to 200, the iteration number was 100, and the size of batch processing in each iteration was 400. LLE was used to perform the dimensionality reduction on the dictionary.

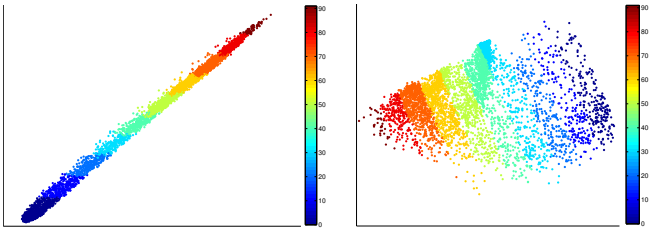
In order to facilitate the visualization of feature vectors in the reduced dimension, the regions were assigned with different colors according to the intensity information. Fig. 6 shows the results of SE and GCSE. Both of them can discriminate the regions correctly. The difference is that GCSE preserves the additional geometric information in the original data, while SE just embedded the features into different areas. For the computing time, GCSE takes around 3.74 s each iteration, and SE requires about 862.93 s for every round. Although the number of iteration for SE is less than GCSE's, SE is time-consuming when the total computing time is taken into account.



(a) k = 50 (b) k = 100

(c) k = 150 (d) k = 200

Fig. 5: Residual errors using different configurations on Toroidal Helix data set.



(a) SE. (b) GCSE.

Fig. 6: Comparison between sparse embedding and proposed method.

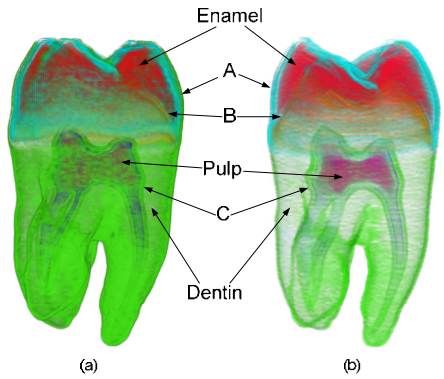


Fig. 7: Rendered images using different transfer functions.

C. Evaluation of Transfer Function in DVR

In this section, the rendering quality of boundaries in the direct volume rendering is mainly evaluated using different transfer functions on human tooth CT. Conventional methods and SE can not handle this data set with the size of 6×1180170. One transfer function (TF-IG) is designed based on the intensity and gradient information, the other one (TF-GCSE) is obtained in the GCSE domain.

The rendering results are shown in Fig. 7. One of the disadvantages for DVR using TF-IG is that some regions close to the boundary can not be isolated exactly. As shown in Fig. 7a, when the background/dentin boundary is labelled with the desired color and opacity, the pulp/dentin and background/enamel boundaries are also colored. Because they belong to the same range defined by the intensity and gradient. In Fig. 7b, TF-GCSE provides a clearer view on the boundaries. Because the Hessian measures are involved in the resultant features obtained by GCSE. Materials involved in the human tooth can also be identified based on TF-GCSE, such as enamel, dentin and pulp.

VI. Conclusions

In this paper, a novel scheme for dimensionality reduction is proposed by encoding the geometrical structure based on a geometry distance. GCSE performs well by learning a good dictionary as a problem of SR. To the best of our knowledge, it is the first time to introduce SR into the design of transfer function in direct volume rendering. More importantly, GCSE enhances the capability to deal with the large data set and enables fast implementation based on stochastic gradient descent. Experimental results show the convergence and robustness of GCSE on the synthetic data. We also have tested GCSE on the CT data containing millions of samples, and the features in the reduced dimension are able to discriminate the materials and boundaries in the volume.

VII. Acknowledgement

This research is partially supported by National Science Foundation of China under Grant 61305114.

References

[1] Jian Wang, Pascal Fallavollita, Lejing Wang, Matthias Kreiser, and Nassir Navab. Augmented reality during angiography: integration of a virtual mirror for improved 2d/3d visualization. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 257–264. IEEE, 2012.

[2] Kevin Cleary and Terry M Peters. Image-guided interventions: technology review and clinical applications. *Annual review of biomedical engineering*, 12:119–42, August 2010.

[3] Gordon Kindlmann. Transfer functions in direct volume rendering: Design, interface, interaction. *Course notes of ACM SIGGRAPH*, 2002.

[4] LJP Van der Maaten, EO Postma, and HJ Van Den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1–41, 2009.

[5] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[6] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[7] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[8] Shivani Agarwal, Aatif Awan, and Dan Roth. Learning to detect objects in images via a sparse, part-based representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(11):1475–1490, 2004.

[9] J-L Starck, Michael Elad, and David L Donoho. Image decomposition via the combination of sparse representations and a variational approach. *Image Processing, IEEE Transactions on*, 14(10):1570–1582, 2005.

[10] Michael Elad. *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer, 2010.

[11] Hien V Nguyen, Vishal M Patel, Nasser M Nasrabadi, and Rama Chellappa. Sparse embedding: a framework for sparsity promoting dimensionality reduction. In *Computer Vision–ECCV 2012*, pages 414–427. Springer, 2012.

[12] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM, 2009.

[13] Carl Meyer. *Matrix analysis and applied linear algebra book and solutions manual*, volume 2. Siam, 2000.

[14] Nils Barth. The gramian and k-volume in n-space: some classical results in linear algebra. *J Young Investig*, 2, 1999.

[15] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591, 2001.

[16] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19:801, 2007.

[17] Simon Perkins and James Theiler. Online feature selection using grafting. In *In International Conference on Machine Learning*, pages 592–599. ACM Press, 2003.

[18] Bradley Efron, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. Least angle regression. *Annals of Statistics*, 32:407–499, 2004.

[19] SA Pawiro, P Markelj, F Pernuš, and C Gendrin. Validation for 2D/3D registration I: A new gold standard data set. *Medical Phys*, 38(3):1481–1490, 2011.

[20] A. Shapiro and Y. Wardi. Convergence analysis of gradient descent stochastic algorithms. *Journal of Optimization Theory and Applications*, 91(2):439–454, 1996.