

## Middlesex University Research Repository:

an open access repository of  
Middlesex University research

<http://eprints.mdx.ac.uk>

Qian, Yu; Hui, Rui; Gao, Xiaohong W., 2012. 3D CBIR with sparse coding for image-guided neurosurgery. Available from Middlesex University's Research Repository.

---

### Copyright:

Middlesex University Research Repository makes the University's research available electronically.

Copyright and moral rights to this work are retained by the author and/or other copyright owners. No part of the work may be sold or exploited commercially in any format or medium without the prior written permission of the copyright holder(s). A copy may be downloaded for personal, non-commercial, research or study without prior permission and without charge. Any use of the work for private study or research must be properly acknowledged with reference to the work's full bibliographic details.

This work may not be reproduced in any format or medium, or extensive quotations taken from it, or its content changed in any way, without first obtaining permission in writing from the copyright holder(s).

If you believe that any material held in the repository infringes copyright law, please contact the Repository Team at Middlesex University via the following email address:  
[eprints@mdx.ac.uk](mailto:eprints@mdx.ac.uk)

The item will be removed from the repository while any claim is being investigated.

# Author's Accepted Manuscript

3D CBIR with sparse coding for image-guided  
neurosurgery

Yu Qian, Rui Hui, Xiaohong Gao



[www.elsevier.com/locate/sigpro](http://www.elsevier.com/locate/sigpro)

PII: S0165-1684(12)00382-9  
DOI: <http://dx.doi.org/10.1016/j.sigpro.2012.10.020>  
Reference: SIGPRO4853

To appear in: *Signal Processing*

Received date: 1 March 2012  
Revised date: 11 October 2012  
Accepted date: 29 October 2012

Cite this article as: Yu Qian, Rui Hui and Xiaohong Gao, 3D CBIR with sparse coding for image-guided neurosurgery, *Signal Processing*, <http://dx.doi.org/10.1016/j.sigpro.2012.10.020>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## 3D CBIR with sparse coding for image-guided neurosurgery

Yu Qian<sup>1</sup>, Rui Hui<sup>1,2</sup>, Xiaohong Gao<sup>1\*</sup>

<sup>1</sup> School of Engineering and Information Sciences, Middlesex University, London, NW4 4BT, UK

<sup>2</sup> Department of Neurosurgery, General Navy Hospital, Beijing, China

**Abstract:** This research takes an application-specific approach to investigate, extend and implement the state of the art in the fields of both visual information retrieval and machine learning, bridging the gap between theoretical models and real world applications. During an image-guided neurosurgery, path planning remains the foremost and hence the most important step to perform an operation and ensures the maximum resection of an intended target and minimum sacrifice of health tissues. In this investigation, the technique of content-based image retrieval (CBIR) coupled with machine learning algorithms are exploited in designing a computer aided path planning system (CAP) to assist junior doctors in planning surgical paths while sustaining the highest precision. Specifically, after evaluation of approaches of sparse coding and K-means in constructing a codebook, the model of sparse codes of 3D SIFT has been furthered and thereafter employed for retrieving. The novelty of this work lies in the fact that not only the existing algorithms for 2D images have been successfully extended into 3D space, leading to promising results, but also the application of CBIR, that is mainly in a research realm, to a clinical sector can be achieved by the integration with machine learning techniques. Comparison with the other four popular existing methods is also conducted, which demonstrates that with the implementation of sparse coding, all methods give better retrieval results than without while constituting the codebook, implying the significant contribution of machine learning techniques.

**Keywords:** CBIR, Computer aided path planning, Neurosurgery, 3D SIFT, Sparse coding

### 1 Introduction

At present, many research fields have developed plethora well-developed theoretical models that are waiting to be applied, whereas each application domain lacks tailor-made specific algorithms that fit for purpose. To fill this gap, this paper focuses on the development of path planning system for image-guided neurosurgery via the application of techniques resulted from both visual information retrieval and machine learning domains.

---

\* Corresponding author: x.gao@mdx.ac.uk. Tel: +44 208 411 2252.

Thanks to the advanced imaging techniques, image-guided key-hole brain surgery has made a significant impact on patients not only by improving clinical outcomes, but also by reducing their recovering time, cost, and psychological issues with little scars. In this regard, the procedures involve path planning based on the acquired patients' MR or CT images, physical drilling of a burr hole, and inserting a probe through the hole [1] to extract the tumour matter. Since the last two steps, i.e., drilling a hole and inserting a probe, follow the path that is planned in the first step, the planning stage plays a crucial part in ensuring a successful clinical outcome of a surgery. As such, the current planning stage has been conducted mainly by experts manually due to its delicate nature of the brain. To alleviate expert's pressure and to train junior doctors, computer aided path planning (CAP) system is being developed.

Although still in a research field, a number of automatic CAP systems have been proposed. For example, it has been reported by a number of researchers that the analogous path can be worked out automatically not only in industrial robotic remit [2] but also in medical treatment. On the other hand, by the application of Java programming, a group of reserchers [3] have developed an autonomous robot motion planning system that can visualize each phase of the planning process graphically. Similarly, another graphical technique to view multimodal images interactively is presented at [4], facilitating complementary tools to allow neurosurgeons to manipulate segments of images, leading to the application of path planning for surgical therapy in epilepsy. In this case, the data that can be visualised simultaneously include modalites of MRI, CT, fMRI and PET. In reality, however, it is very expensive to have more than two modality data acquired for each patient. In addition, a method of automatic neurosurgical path searching is described at [5], which integrates the structure of blood vesssels obtained from MRA and brain tissues. Their path searching machanism consists of a voting scheme by assiging each region assigned with a number between 1 to 4 according to the importance of the region; i.e., a smaller number can be given to a region that is further from blood vessels, by which the regions with minimum values are classified as the safest paths. Although this approach has a potential to provide a comprehensive system to help neurosurgeons to design a perfect path, the classification of the importance values is not so specific and MRA is too extravagant for ordinary hospitals and average patients. Therefore, the search for neurosurgical path planning in this investigation focuses on the reasoning of those successful cases that have already been collected by the application of technique of content-based images retrieval (CBIR).

In general, a CBIR system extracts features of images in terms of their global visual information, such as colour, texture, and shape, and then represents these features using mathematical vectors that in turn are employed to index each image. When a query image is submitted, the system only needs to extract these features from the query and performs the comparison with the feature database that has been stored in advance. In this way, the retrieval process of an image can be as fast as that in a text-based system since the similarity calculation is based upon numerical data only. In the past two decades or so, CBIR has been researched mainly on two dimensional images [6]. Only recently, CBIR for 3D images have been attempted with very promising results [7]. Because of the subjectivity nature of visual information and the differences between users' search intention, CBIR remains within a research domain.

Within CBIR, machine learning techniques have been introduced into in two ways in an attempt to further improve the its effectiveness. One way is to adopt unsupervised learning technique of Bag of Words (BoW) paradigm to train a codebook of the visual features of a training dataset which was first introduced in [8]. As a result, an image is represented by using the statistic summaries of the appearance of each word in the codebook as feature vectors that are then in turn utilized in image retrieval and classification. Another approach is to apply supervised learning technique by the employment of the technique of relevance feedback [9,10], leading to the study of both positive and negative examples selected by users from retrieved results, to discover and capture users' real intention, and to the modification of the retrieval process, thus obtaining retrieval results entailing the user's actual request as precise as possible.

In this paper, 3D CBIR is to be applied in assisting path planning procedures for image-guided neurosurgery, which is coupled with a machine learning technique of sparse coding, incorporating both experts' experience and knowledge by way of learning from the past through the reasoning of the existing data.

Figure 1 demonstrates two examples of planned paths employed during two operations, which are initiated by an expert, where the lines can be viewed from different directions indicating the paths that a probe (i.e., a catheter) has been inserted through.

**Fig.1.** Two examples of surgical paths that are indicated using lines.

At present, the main factors attributed to the definition of a path for an experienced physician is the location, size, and shape of a lesion, whereby, lesions of similar characteristics would result in a similar surgical path. Although a patient's pathology also affects the path planning, it varies between patients. Hence, in this study, this element is not considered for the time being. It is expected that by giving the current patient's MR images, the developed CAP system should return back the past cases with similar lesions in terms of location, size, and shape; thereafter their accompanying surgical paths that will be utilised to guide the planning of the current case of interest.

Towards this end, in this investigation, a 3D brain codebook is constructed by using the paradigm of Bag of visual Words (BoW) [8], 3D Scale Invariant Feature Transform (3D SIFT), and Sparse coding. Sparse coding originated to extend the traditional classifier in controlling over-fitting problems [11] that arose in support vector machines (SVM). Since then, this idea has been applied in many fields in enforcing equal treatment of scattered data [12]. In this paper, the MR images of a 3D brain are firstly specified using 3D SIFT sparse codes, whereas the representation of potential volume of interests (VOIs) adopts the approach of spatial max-pooling of 3D sparse codes. While performing retrieval, the similarity measurement is only carried out on the VOIs between a query and its corresponding brain regions in the database, expediting the processing. In doing so, a pre-processing stage of a selection of potential VOIs is introduced into querying stage via statistical analysis of bilateral symmetry of a query brain.

The remainder of the paper is structured as follows. Section 2 starts to analyze the patterns of surgical path that precipitates the following work addressed at Section 3 that details the methodology furthered in this investigation. Section 4 displays the experimental results, whereas conclusion and discussion are then drawn in Section 5.

## **2. Analysis of surgical path**

Based on the existing data, preliminary study has shown that similar tumour locations will result in similar surgical paths, which can be demonstrated in the following experiment.

One hundred patients' data of both MR images and corresponding surgical paths that have been designed to perform the operations for tumour removal are analyzed. As illustrated in Figure 1, two examples of path patterns are shown, i.e., right and left, with reference to the sagittal plane passing through the nose to be illustrated below.

Because of this collection of data contains mainly the type of tumour of craniopharyngioma that occurs in the region of parasellar, near pituitary gland as shown in Figure 1. This type of tumour is usually accompanied by a cyst of liquid/semi-liquid form, which is the reason why image-guide key-hole surgery being the most common and popular treatment. Since the region of parasellar occupies a specific area in the brain, the classification of the path pattern is carried out within two categories for the purpose of illustration, which are on the left and on the right hand side of the brain with reference to the middle plain (sagittal) as depicted in Figure 1, where those lines are the indication of surgical paths. According to the location of tumours and the surgical paths, manual measurements are taken with reference to the most posterior point of the fourth ventricle at the midline, the point that usually retain its position even at the presence of tumours [13]. Figure 2 demonstrates the centre positions of one hundred cases at XY plane. The axial line where  $x=0$  represents the sagittal midline projected onto the XY plane. Likewise, the one hundred operation path lines are plotted in Figure 3 in terms of line angles. Again the midline forms part of the sagittal plane. The lines are expressed using angles. With only 8 exceptions, the locations of the centre points on the left of the midline in Figure 2 correspond to the right group positioned in Figure 3. Similarly, the tumours located on the right in Figure 2 will normally have their operation path falls in the left category of the clustered data shown in Figure 3. Therefore it can be concluded that similar tumour location will derive similar surgical path, the assumption that the following work is based upon. The detailed path classification with corresponding to tumour location forms the main content of a separate paper that is under writing.

**Fig.2.** The locations of tumour centres of one hundred in the XY plane, in which the line  $x=0$  represents the sagittal midline pass through the 4<sup>th</sup> ventricular point, the internal reference point.

**Fig.3.** The angles of the 100 path lines projected onto XY and YZ planes. The line  $x=0$  indicates the sagittal midline.

### 3. Methodology

The major contribution of this work is to further those well-established algorithms for 2D images into 3D space for CBIR by the inclusion of machine learning techniques, which includes the construction of codebook and retrieval of 3D images, leading to a potential clinical application in the field of computer aided neurosurgery.

#### 3.1 The creation of 3D brain codebook

Figure 4 schematically illustrates the procedures of creating a visual dictionary of brain images, i.e., a codebook. The key phases follow the paradigm of bag of words (BoW), which constitutes the setup of visual feature descriptors using 3D SIFT, the generation of visual dictionary via sparse coding (left graph) and the image representation based on visual words (right graph). The BoW paradigm transforms images into a set of ‘visual vocabulary’ and represents them using the statistics data of the appearance of each word as feature vectors, which is composed of four phases to be detailed below.

**Fig.4.** A flowchart of dictionary construction and image representations.

##### **Phase 1: Pre-processing --- Spatial normalization**

By nature, each brain image varies in both shape and size, in particular for those brains with lesions. In order to make inter-brain comparable, it is necessary to transform each individual brain into a standard template.

In this regard., statistical parametric mapping (SPM5) [14] is used to spatially normalize a brain image to an MNI template [15] in ensure that all brain images are with the same size of  $157 \times 189 \times 69$  voxels.

##### **Phase 2: local feature extraction ---3D SIFT descriptors**

SIFT descriptors [16], on the other hand, are invariant to location, scale and rotation, and provide robust feature matching across a substantial range of changes in illumination while with the presence of noise. It is therefore widely applied in the domain of object recognition

and image stitching. In addition, a 3D extension of the SIFT algorithm has recently been proposed in the literature for 3D volumetric data analysis, such as, action recognition in video volumes [17], object recognition in CT complex volumes [18], 3D medical registration and panoramic medical image stitching [19].

Since in our work, the main intended targets, i.e., tumour, are usually known, the extraction of features constitutes a number of simple and straightforward steps. For example, in order to describe local features from different parts of a brain, a normalized 3D volumetric brain is divided into 512 non-overlapping equally sized sub-volumes, giving rise to 8 sub-volumes along each of  $x$ ,  $y$ ,  $z$  axes respectively. Subsequently, 3D SIFT descriptors are applied to extract local features by computing a 3D gradient orientation histogram within each sub-volume.

As illustrated in Figure 5(b), each sub-volume of a brain is divided into  $2 \times 2 \times 2 = 8$  sub-blocks, within which the magnitude and orientation of the gradient are calculated. For example, for each sub-blocks, the magnitude of the gradient is accumulated to the corresponding bin of the gradient orientation. As a result, the bin of 3D gradient orientation is approximated with a mesh of small pieces of a 3D volume as illustrated by a triangle in Figure 5(d) using tessellation techniques. Therefore, the gradient orientations pointing to one triangle belong to the same bin, as marked in black points in Figure 5(d). In this way, the total number of bins is calculated as  $N \times 4^{\text{Tessellation\_level}}$ ; where  $N$  indicates the original number of the mesh in a sphere, and  $\text{Tessellation\_level}$  representing the recursive times to subdivide the input mesh by using triangular quadrisection. At present,  $N$  is set to 20 whereas  $\text{Tessellation\_level}$  is set to 1, leading to 80 bins of gradient orientation in the 3D space. Thereafter, each sub-block is accumulated into its own sub-histogram. As a result, the 3D SIFT descriptor  $X$  of each sub-volume has of 640 dimensions ( $=2 \times 2 \times 2 \times 80$ ).

**Fig.5.** 3D SIFT descriptors.

### **Phase 3: Visual vocabulary construction---Sparse coding**

Once the 3D SIFT features, i.e., the candidates for unit elements, or “words” in a visual dictionary, are extracted from each sub-volume, sparse coding follows.

Sparse coding [20, 21, 22] models data vector as a sparse linear combination of a set of basic elements known as a dictionary encodes each descriptor of an image by solving the optimization problem as formulated in Eq.(1).

$$\min_{U,V} \sum_{m=1}^M \|x_m - Vu_m\|_2^2 + \lambda \|u_m\|_1 \quad \text{Subject to: } \|v_i\| \leq 1, \quad \forall i = 1, \dots, K \quad (1)$$

where  $X = [x_1, x_2, \dots, x_M]$  ( $x_m \in R^{dx1}$ ) refers to a set of 3D SIFT descriptors as described above from a 3D training dataset;  $V = [v_1, v_2, \dots, v_K]$  ( $v_i \in R^{dx1}$ ) indicates the  $K$  bases, also known as the dictionary or codebook; and  $U = [u_1, u_2, \dots, u_M]$  ( $u_m \in R^{Kx1}$ ) denotes sparse codes for images based on codebook  $V$ .  $M$  refers to the number of training samples in the training dataset.

In the training stage, 3D descriptors extracted from random sub-volumes obtained from a training dataset are applied to off-line training on the codebook  $V$  by solving Eq.(1) using alternating optimizing over either  $V$  or  $U$  while fixing the others. When fixing  $V$ , feature-sign search is adopted by optimizing over each  $u_i$ . When fixing  $U$ , the Lagrange dual is adopted by optimizing over  $V$ . The details of the optimization process are similar to that described at [20].

#### Phase 4: 3D brain image representations---Sparse codes

In the coding stage, 3D SIFT descriptors  $x_i$  extracted from each sub-volume of a normalized 3D brain can be encoded as  $u_i$  by inputting the trained codebook  $V$  in Eq.(1), which leads to a 3D brain being represented as a set of  $\bar{U} = [u_1, u_2, \dots, u_N]$ , where  $N$  is the total number of the sub-volumes in an 3D brain.

On the other hand, on the assumption that each sub-volume segmented from the same location of a brain constitutes similar tissue structure (after normalization), sub-codebook is formed for each part of the brain. In total, 64 sub-codebooks are generated from 64 ( $=4 \times 4 \times 4$ ) non-overlapping equally sized partial brain respectively. Since 3D SIFT descriptors had been extracted from 512 sub-volumes of each brain, eight 3D SIFT descriptors will attribute to the

corresponding sub-codebooks. Based upon 120 MR images in our collection, 960 ( $=120 \times 8$ ) features are employed to represent each part of the brain. Consequently, in the training stage, 500 descriptors randomly selected from these 960 sub-volumes are applied to train the sub-codebook  $V$ , where the size of  $K$  for each  $V$  is set to be 30. Hence, a 3D brain codebook is composed of 64 sub-codebooks. Subsequently, sparse codes for each sub-volume will adopt the corresponding sub-codebooks.

### 3.2 Lesion-based image retrieval

Figure 6 illustrates the proposed framework for retrieval. Firstly images undergo spatial normalization with reference to the MNI template via SPM5. Then those normalized volumetric data are divided into 512 non-overlapping equally sized sub-volumes, upon which 3D SIFT sparse codes are calculated based on the trained codebook, leading to the creation of a 3D brain feature database. On the other hand, at querying stage, the potential VOI selection is introduced after spatial normalization. Therefore, 3D SIFT sparse codes of a query are extracted only from these potential VOIs using spatial max-pooling function forming feature vectors that are then compared with the corresponding features in the feature database to obtain retrieved results.

**Fig.6.** Schematics diagram of proposed content-based 3D brain retrieval.

#### 3.2.1 Potential VOI selection

The main purpose of this study is to search images with lesions of similar location, size or shape. Although the feature database has been implemented in advance, the procedure of processing a query should be conducted in real time. In another word, after a query being submitted, 3D SIFT sparse codes should be calculated from its 512 sub-volumetric regions together with the calculation of similarity distances, within restricted sub-second time period. To this end, while maintaining the overall performance of retrieval, the detection of potential lesions from sub-volumes is carried out first to highlight the abnormalities, such as tumours, to speed up the retrieving process.

In doing so, the characteristic of bilateral symmetry of a brain along its sagittal plane is assumed. By comparing the left half with its right counterpart of a hemisphere along this

middle symmetry plane, the abnormality is envisaged to be singled out. The details of potential VOI selection by using Bhattacharya Coefficient and experimental results can be found in [23, 24].

### 3.2.2 VOI representations--- Spatial max-pooling

After the affirmation of VOI from a query is established, the 3D SIFT sparse codes are calculated exclusively from each sub-volume within this VOI of the query using the corresponding sub-codebook  $V$ . The VOI representations  $Z = \{z_i, i = 1, 2, \dots, K\}$  is computed by a max-pooling function expressed in Eq.(2).

$$z_i = \max \left\{ |u_{ij}|, j = 1, 2, \dots, S \right\} \quad (2)$$

where  $u_{ij}$  in Eq.(2) indicates the  $i^{\text{th}}$  ( $i \in [1, K]$ ) space code for the  $j^{\text{th}}$  ( $j \in [1, S]$ ) sub-volume.  $K$  indicates the size of the corresponding sub-codebook  $V$ , and equals to 30. Whereas  $S$  refers to the total number of the sub-volumes within a VOI, and is set to 8 in this study.

A VOI representations  $Z$  of a query are subsequently compared with the corresponding representations summarized from the counterpart of sub-volumes of images in the code database in an attempt to search brains with similar lesions.

### 3.2.3 Similarity measurement

The histogram intersection and Chi-square histogram are applied to measure the degree of similarity between two brains represented as  $Z_Q$  and  $Z_P$ , as given in Eqs.(3) and (4) respectively.

$$D(Z_Q, Z_P) = \sum_{i=1}^K \min(z_{Qi}, z_{Pi}) \quad (3)$$

$$\chi^2(Z_Q, Z_P) = \frac{1}{2} \sum_{i=1}^K \frac{(z_{Qi} - z_{Pi})^2}{(z_{Qi} + z_{Pi})} \quad (4)$$

For histogram intersection distance expressed in Eq.(3), the more similar between a query ( $Q$ ) and a brain ( $P$ ) is, the bigger value the  $D$  remains . Therefore, the retrieved results are ranked in descending order based on the value of  $D$ . By contrast, the retrieval results are ranked in ascending order using Chi-square histogram distance formulated in Eq.(4). The evaluation results using these two distance formulae are given in Section 4.3.

## 4. Experimental results

### 4.1 Dataset

3D MR brain images (n=120) including both normal (34) and lesioned (86) data, are collected from Neuro-surgery Centre at Beijing General Navy Hospital, China, and are utilized to evaluate the proposed approach. These images are in DICOM (Digital Imaging and Communications in Medicine) format with 16 bit gray-level resolution. The average resolution of these images is  $500 \times 500 \times 45 \text{ mm}^3$ . As detailed in Table 1, the ground truth data based on the location of lesions is created by clinicians in the Neuro-surgery Centre. Preliminary result as demonstrated in Section 2 has shown that similar lesion locations will arrive at similar surgical paths with the accuracy of over 90%.

Supposedly, each normalized brain is divided into 8 non-overlapping equally sized sub-volumes and labelled from 1 to 8 in anticlockwise and front-to-back order. In Table 1, the first row is the labeling number of the location of lesions e.g., ‘1’ refers to the lesion located in the front top left part of the brain, whilst ‘9’ in last column refers to normal case. The second row is the total number of brains containing lesion in such regions in the database. Some brains have more than one lesion scattered in different part of the brain.

### 4.2 Sparse coding vs Kmeans for the construction of 3D brain codebook

In this study, the approach of sparse coding is selected to ensure the optimal selection in the generation of codebook and image retrieval. Evaluation of both sparse coding and K-means is therefore carried out first.

In terms of image retrieval, two aspects are concerned by using the well-known approaches of average quantization error (AQE) and similarity correspondence. In doing so, each normalized brain is divided into 512(=8×8×8) non-overlapping equally sized sub-volumes, and 5000 3D SIFT descriptors are extracted randomly from the sub-volumes of 120 images. Both K-means and sparse coding are then applied to train the 3D brain codebook with the size of 128. We initialize the 128 bases codebook randomly and run each method. After 30 iterations, the convergence curves for K-means and sparse coding are shown in Figure 7 that shows that both methods converge after the third iteration.

**Fig.7.** Convergence of K-means and sparse coding for training 3D brain codebook.

#### 4.2.1 Average Quantization Error (AQE)

Average Quantization Error (AQE) is an important value for evaluating the quality of feature quantization, and is formulated as Eq.(5) for sparse coding and Eq.(6) for K-means.

$$\text{AQE\_Sparse-coding} = \frac{1}{N} \sum_{n=1}^N \|x_n - Vu_n\|^2 \quad (5)$$

$$\text{AQE\_K-means} = \frac{1}{N} \sum_{n=1}^N \|x_n - C_n\|^2 \quad (6)$$

where  $X = [x_1, x_2, \dots, x_N]$  ( $x_n \in R^{d \times 1}$ ) refers to a set of 3D SIFT descriptors from the dataset, and  $U$  denotes their sparse codes based on a trained codebook  $V$ . In addition,  $C$  is a set of trained cluster centre (codebook) by using K-means method and  $C_n$  represents a cluster centre of  $x_n$  assigned by using KNN.

With regard to the result, smaller quantization error indicates less information loss in the feature quantization, which will enhance the retrieval performance. Based on the dataset of 120 brain images, i.e.,  $N = 61440$  (i.e., 512(=8×8×8 sub-volumes) ×120), the AQE values for sparse coding and K-means are given in Table 2. In comparison with K-means, sparse coding can decrease quantization loss by 0.17, which is due to the fact that sparse codes preserve the spatial relation of 3D SIFT features by using sparse coefficients  $u_n$  to combine the codebook  $V$ , whilst K-means does not preserve these relation among similar SIFT features.

#### 4.2.2 Similarity correspondence

The comparison between the two approaches is furthered into the calculation of similarity correspondence. The main aim of the codebook training is to obtain the best image representation by using BoW paradigm (in Section 3.1), and further improve retrieval results. In CBIR, good representations suggest to preserve the similarity of original features. Towards this end, from the aforementioned 120 datasets, each 3D brain is divided into 64 ( $=4 \times 4 \times 4$ ) non-overlapping sub-volumes, with 200 pairs of sub-volumes being randomly selected. Based on the BoW paradigm, a sub-volume can be represented using the statistics of the appearance of each word of codebook. In this experiment, the histogram of visual words from K-means codebook and spatial max-pooling (in Section 3.2.2) of sparse codes are employed to represent each sub-volume respectively. Then the measures of the tendency between similarities calculated for each approach is under way, including pairwise similarity of 3D SIFT features and the similarity of their sparse codes and K-means representations by calculating histogram intersection (formulated in Eq.(3)) respectively. Figure 8 plots their similarity correspondence with their best-fit regression lines. The  $R$  values are 0.75 for sparse coding (in Figure 8 (a)) and 0.52 for K-means (in Figure 8(b)) respectively. The figures clearly depict an evident linear trend of the similarity between sparse codes against the similarity between 3D SIFT features, which justifies our choice of sparse coding for it preserves better similarity of original features than the approach of K-means.

**Fig.8.** The similarity correspondence.

Comparing with K-means in term of AQE and similarity correspondence, sparse coding is selected to train and represent the codebook of our 3D data in this study as the result of the above analysis, i.e., sparse codes preserve the spatial relation of visual features by assigning each feature to a number of visual words, while K-means causes severe information loss by assigning each visual feature to only one visual word, especially for those features located at the boundary

### 4.3 Retrieval results

The other four well-established methods employed in CBIR are also exploited in this investigation with the extension to 3D. They are of texture representations including 3D Grey Level Co-occurrence Matrices (3D GLCM), 3D Wavelet Transforms (3D WT), 3D Gabor Transforms (3D GT) and 3D Local Binary Pattern (3D LBP). The details of these 3D texture representations for 3D brain can be found in [23, 24, 25].

In this section, the performance of retrieval with and without sparse coding by using the aforementioned five methods (3D SIFT and four texture representations) is also evaluated respectively. Table 3 shows the value of Mean Average Precision (MAP) for ten queries across the whole datasets with and without sparse coding by using these five methods. The retrieval results with sparse coding using histogram intersection and Chi-square distance are listed in column 3 and 4 respectively. Table 3 demonstrates the approach of 3D SIFT that has been furthered in this research outperforms the other four with the average MAP of 0.4098 using Chi-square histogram distance. In addition, the implementation of sparse coding improves the performance of all five approaches, specifically for 3D GT, implying the significance of contribution of machine learning technique. Due to the limited size of the collected database ( $n=120$ ), considerable improvement is not expected. However, the trend is very promising.

Correspondingly, Figures 8 illustrates the average Precision Recall Graph at  $[0, 0.5]$  recall rate in comparison with Figure 9(a) and (b) depicting the results with sparse coding using histogram intersection and Chi-square histogram distance respectively, and Figure 9(c) displaying the results without sparse coding. Finally, Figure 9(d) shows the Mean of average Precision Recall Graph for all five feature descriptions without sparse coding, with sparse coding either applying intersection histogram or utilizing Chi-square histogram distance.

**Fig. 9.** Average Precision Recall Graph for ten queries.

#### 4.4 Query time

Table 4 lists the averaged query time, including 3D feature extraction, feature representation using sparse coding for the query image together with the retrieving time through the whole database. The second column is the average query time without VOI selection while the third column is the average query time with VOI selection. All methods run in Matlab R2011b with CPU of Inter P8600 1.58GHz and 3.45GB RAM.

In Table 4, the querying time with VOI selection is around 4 times faster than the retrieval without VOI selection. The querying time for 3D GT takes longer than the other methods with 21 minutes, due to the employment of  $128(=4(\text{scales}) \times 16(\text{orientations}))$  times of 3D convolutions for each sub-volume. It appears 3D LBP with sparse coding performs the fastest. This is because that LBP in essence remains a 2D approach by placing three orthogonal planes centered at the centre of each sub-volume to represent this cube, which missed the rest of the contents of the cube apart from that on those three 2D planes, implying larger sub-volumes losing more information.

#### 5. Conclusion and discussion

In this investigation, a suite of well-established models and algorithms are tailored, extended and implemented in a system of path searching for image-guided neurosurgery. Specifically, the technique of 3D CBIR coupled with sparse coding has demonstrated its promising performance. In comparison with the other four popular approaches that are applied in the field of CBIR, the proposed method also shows better results in terms of precision and recall values. However, when it comes to the query time, it takes the approach of 3D SIFT over a minute to complete the retrieval, much slower than three of the other four approaches. Since this retrieval task is for the planning stage while conducting an image-guided neurosurgery, real-time retrieval is not a prerequisite. Nevertheless, a number of improvements can be made in the future. First of all, the sampling size should be increased in an attempt to cover variety of lesion types. At present, nearly all the lesion types belong to tumour, whereas similar operations on head injury and stroke are also performed. Secondly, although each brain has been spatially normalized, due to the distortions caused by the location of lesions, each sub-volume cannot always be described precisely by the vocabularies that are stored in the

corresponding sub-codebooks. Optimal size of each sub-volume should therefore be found out in the future to ensure its comprehensive representations by the codebook. At the moment, the location of a lesion appears to be the biggest factor in the planning stage of a path by clinicians, and therefore constitutes the main content in our retrieval system. In the future, the characteristics of both shape and size of a lesion will also be taken into consideration since they affect the surrounding tissue structure appreciably, and thereafter the codebook that describes each part of the brain. In conclusion, this research showcases an application of theoretical models into real world tasks. For the time being, the developed system in assisting path planning is mainly used for training junior surgeons but it positively exhibits potentials in the clinical applications in the near future, by complementing the existing manual path planning systems.

### **Acknowledgment**

This research is financially funded by both UK JISC and EC under FP7 People programme. Their support is gratefully acknowledged.

### **References**

- [1] Z. Tian, W. Lu, Q. Zhao, X. Yu, S. Qi, R. Wang, From Frame to Frameless Stereotactic Operation—Clinical Application of 2011 cases, *Lecture Notes in Computer Science*, 4987, 2007, pp.18-24.
- [2] J. Vals-Miro, A.S. White A.S., Quasi-optimal trajectory planning and control of a CRS A251 industrial robot, *J Systems and Control, Proc. Institution of Mechanical Engineers (IMechE)*, 216, 2002, pp.343-356.
- [3] A. Elnagar, L. Lulu , A global path planning Java-based system for autonomous mobile robots, *Science of Computer Programming*, 53, 2004, pp.107–122.
- [4] A. Joshi, D. Scheinost, K.P. Vives, D.D. Spencer, L.H. Staib, X. Papademetris, Novel interaction techniques for neurosurgical planning and stereotactic navigation, *IEEE Transactions on Visualization and Computer Graphics(TVCG)* 14 (6), 2008, pp.1587-1594.

- [5] T. Fuji, H. Emoto, N. Sugou, T. Mito, I. Shibata, NeuroPath planner-automatic path searching for neurosurgery, International Congress Series 1256, 2003, pp.587-596.
- [6] M. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, ACM Transactions on Multimedia Computing, Communications, and Applications, 2006, pp. 1-19.
- [7] X.W. Gao, Y. Qian, R. Hui, The state of the art of medical imaging technology: from creation to archive and back, The Open Medical Informatics Journal, 2011, 5 (1) pp.73-85.
- [8] J. Sivic, A. Zisserman. Video Google: A text retrieval approach to object matching in videos, IEEE Conference on Computer Vision (ICCV), 2003, pp. 1470-1477.
- [9] D.Tao, X.Tang, X.Li, X. Wu, Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI). 28(7): 2006, pp.1088-1099.
- [10] D.Tao, X.Tang, X.Li, Y. Rui, Direct kernel biased discriminant Analysis: A new content-based Image retrieval relevance feedback Algorithm. IEEE Transactions on Multimedia 8(4): 2006, pp. 716-727.
- [11] O. Yamashita, M. Sato, T. Yoshioka, F. Tong, Y. Kamitani, Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. NeuroImage 42, 2008, pp.1414-1429.
- [12] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1794-1801.
- [13] R. Pierson, P.W. Corson, L.L. Sears, D. Alicata, V. Magnotta, D. O'Leary, N.C. Andreasen, Manual and Semiautomated Measurement of Cerebellar Subregions on MR Neuromages, 17 (1), 2002, pp. 61-76.
- [14] <http://www.fil.ion.ucl.ac.uk/spm/>. Retrieved in January 2012.

- [15] Montreal Neurological Institute, <http://www.mni.mcgill.ca/>. Retrieved in January 2012.
- [16] D.G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision (IJCV)*, 60 (2), 2004, pp. 91-110.
- [17] P. Scovanner, S. Ali, M. Shah, A 3-Dimensional SIFT Descriptor and Its Application to Action Recognition. In: *ACM Conference on Multimedia*, 2007, pp. 357-360.
- [18] G. Flitton, T. Breckon, N. Megherbi, Object recognition using 3D SIFT in complex CT volumes, In: *British Machine Vision Conference (BMVC)*, 2010, pp.1-12.
- [19] D. Ni, Y. Qi, X. Yang, Y. Chui, T. Wong, S. Ho, P. Heng, Volumetric ultrasound panorama based on 3D SIFT. In: *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2008, pp. 52-60.
- [20] H. Lee, A. Battle, R. Raina, A. Y. Ng, Efficient sparse coding algorithms, *Advances in Neural Information Processing Systems (NIPS)*, 2006, pp.801--808.
- [21] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 1794 – 1801.
- [22] S. Cao, I. Tsang, L. Chia, P. Zhao, Local feature are not lonely – Laplacian sparse coding for image classification, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 3555 – 3561.
- [23] Y. Qian, X. W. Gao, M. Loomes, R. Comley, B. Barn, R. Hui, Z. Tian, Content based image retrieval of 3D images, *The Third International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED 2011*, 2011, IARIA, XPS Press, pp7-12.
- [24] X.W. Gao, Y. Qian, M. Loomes, R. Comley, B. Barn, A. Chapman, J. Rix, R.Hui, Z. Tian, Retrieval of 3D Medical Images via Their Texture Features, *International Journal on Advance in Software*, vol 4 no 3 & 4, 2011.

[25] X.W. Gao, Y. Qian, M. Loomes, R. Comley, B. Barn, A. Chapman, J. Rix, Texture-based 3D image retrieval for medical applications, IADIS e-Health2010, 2010, pp. 29-31.

Accepted manuscript

**Table 1.** Dataset

Location of lesions	1	2	3	4	5	6	7	8	9
Number of brains	14	36	14	28	12	6	8	4	34

Accepted manuscript

**Table 2. Average Quantization Error**

	K-means	Sparse coding
Average Quantity Error(AQE)	0.42	0.25

Accepted manuscript

**Table 3.** Value of Mean Average Precision (MAP).

Methods	Without sparse coding	With sparse coding	
		Histogram intersection	Chi-squared histogram
<b>3D GLCM</b>	0.3034	0.3291	0.3510
<b>3D WT</b>	0.3096	0.3375	0.3687
<b>3D GT</b>	0.3074	0.3863	0.3954
<b>3D LBP</b>	0.3308	0.4027	0.4012
<b>3D SIFT</b>	0.3959	0.4013	0.4098

Accepted manuscript

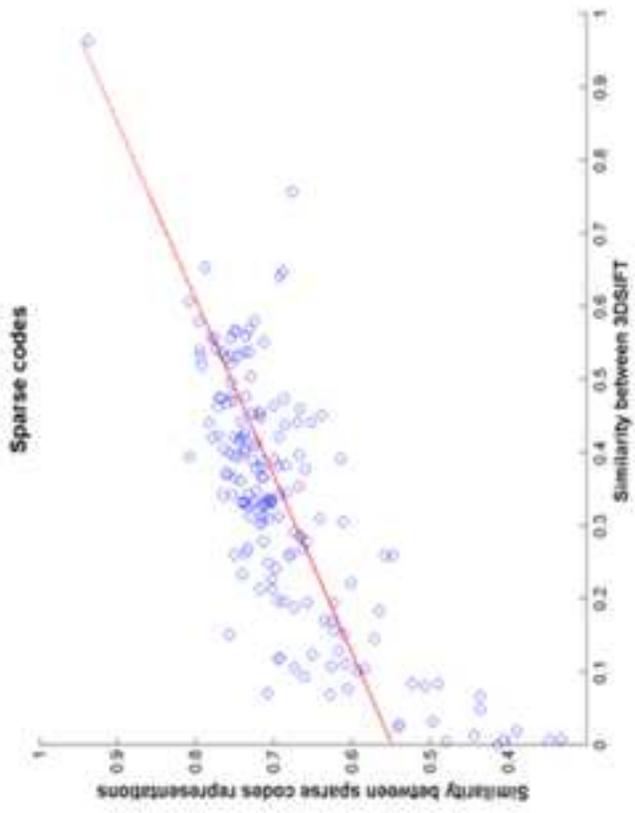
**Table 4.** Query time.

<b>Methods</b>	<b>Without VOI selection</b>	<b>With VOI selection</b>
<b>3D GLCM+SC</b>	1.4047m	0.3612m
<b>3D WT+SC</b>	0.2702m	0.0676m
<b>3D GT+SC</b>	21.3603m	5.6374m
<b>3D LBP+SC</b>	0.1459m	0.04131m
<b>3D SIFT+SC</b>	4.9983m	1.3043m

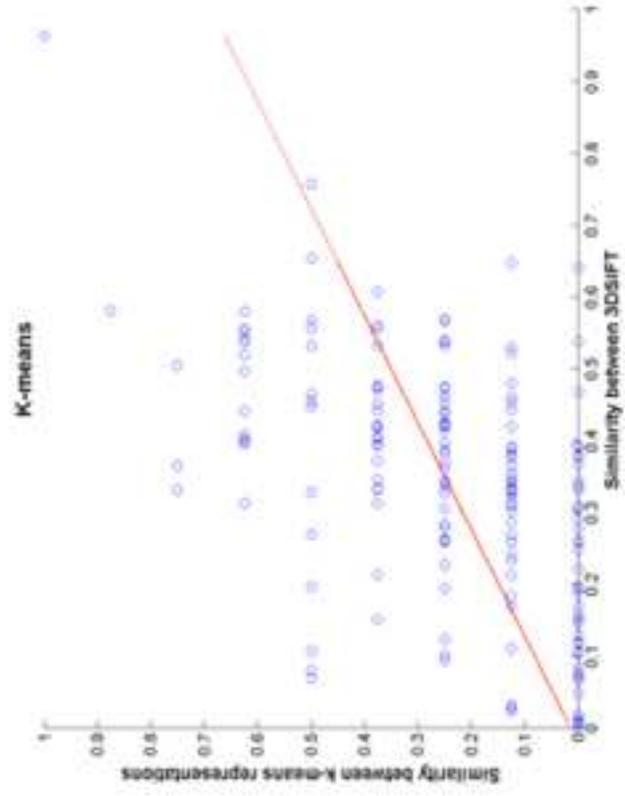
Accepted manuscript

- A codebook for 3D MR brain images has been generated to represent content features
- 3D SIFT coupled with sparse codes are tuned for the creation of a codebook
- Extensions into 3D from 2D algorithms have been refined to fit for purpose.
- The proposed feature representation approach using codebook performs the best.
- The specified research demonstrates potentials in future clinical applications.

Accepted manuscript

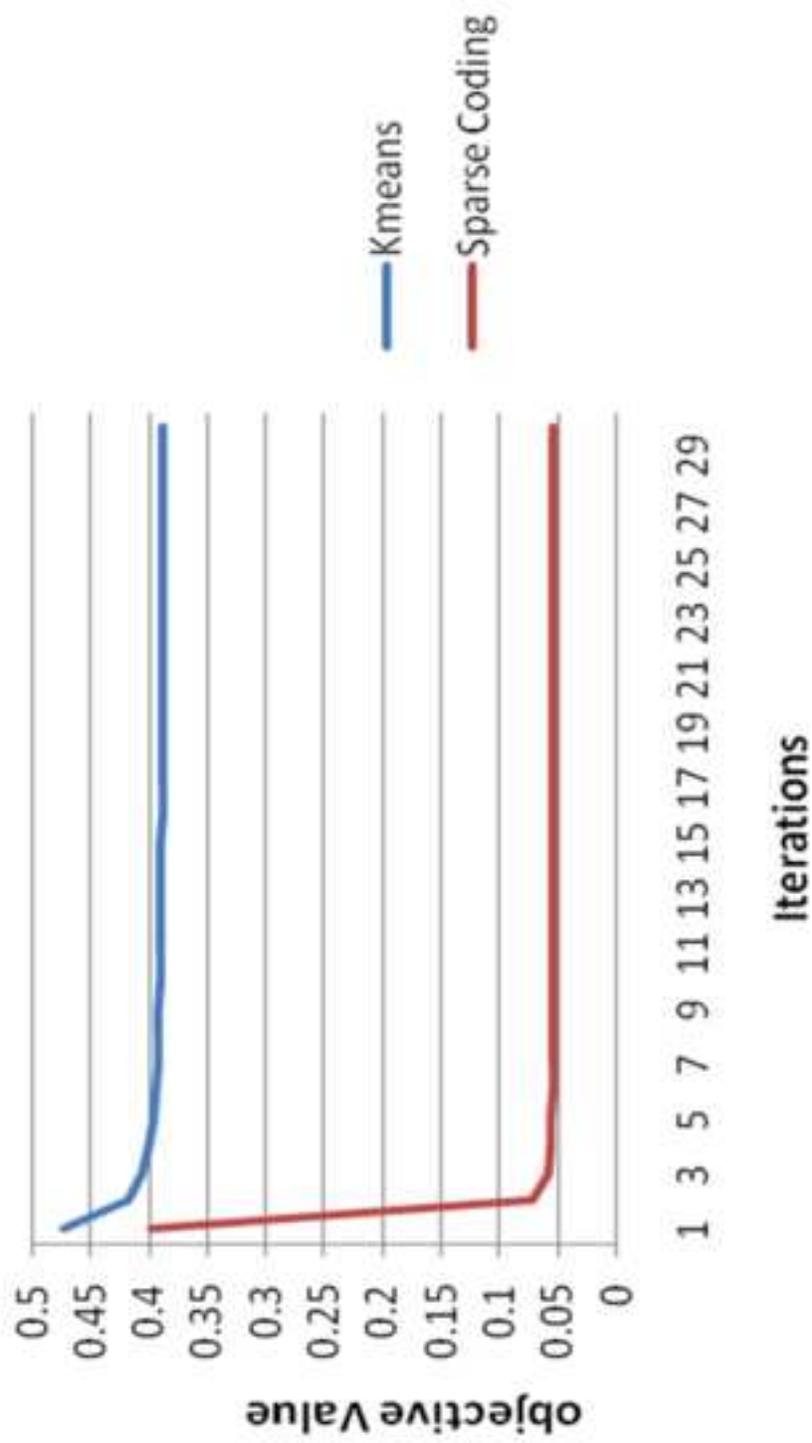


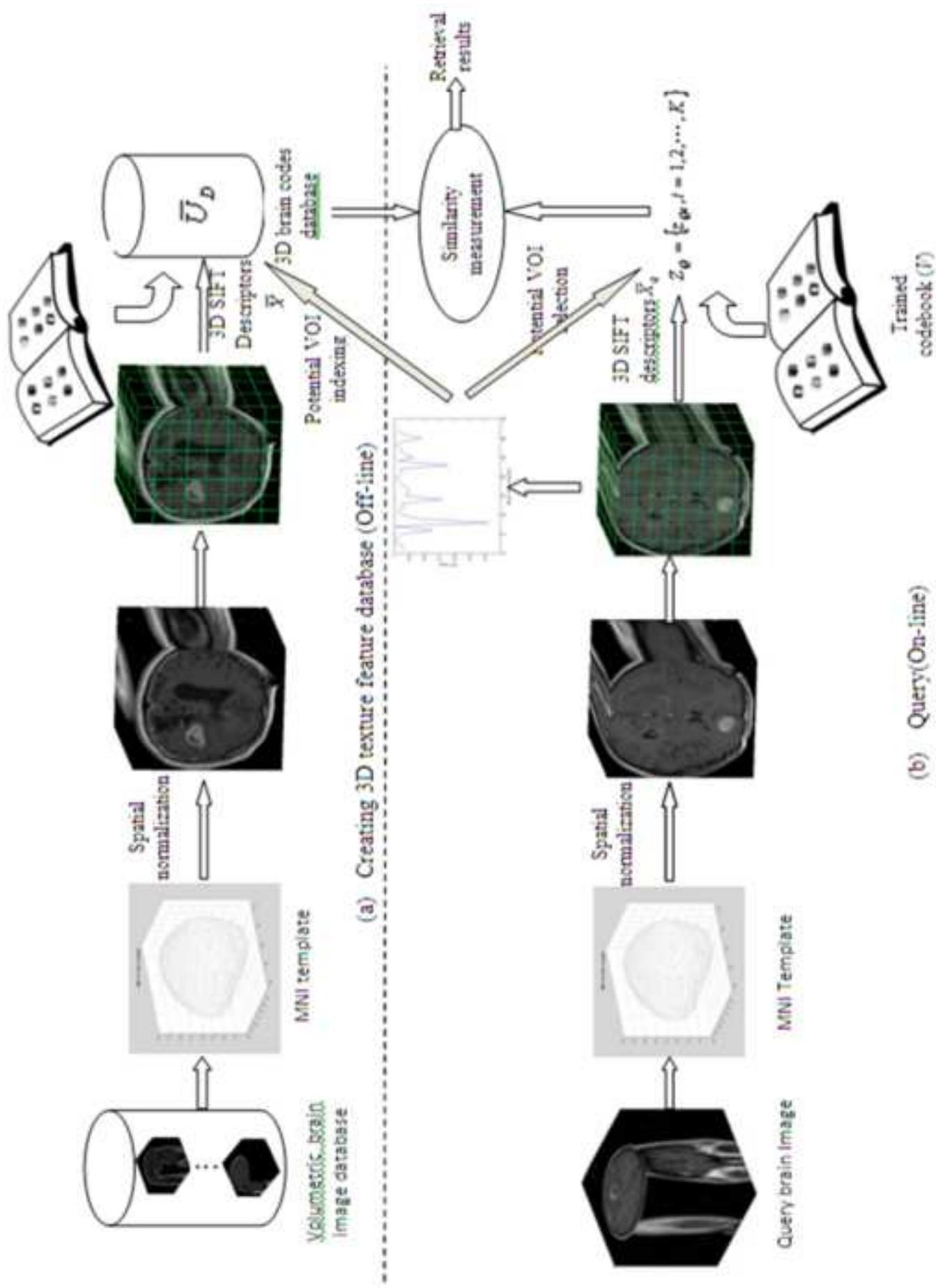
(b) Sparse coding



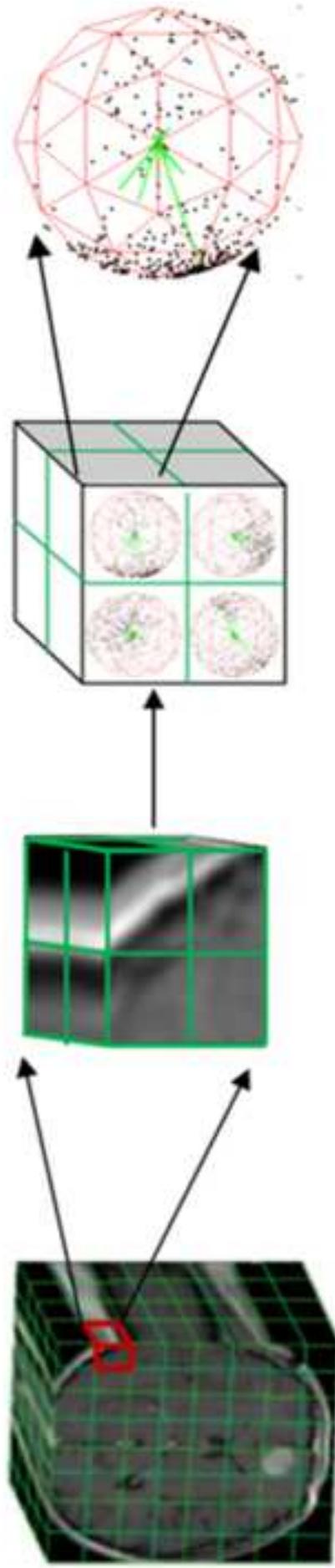
(a) K-means

## Comparison of Convergence





(b) Query(On-line)

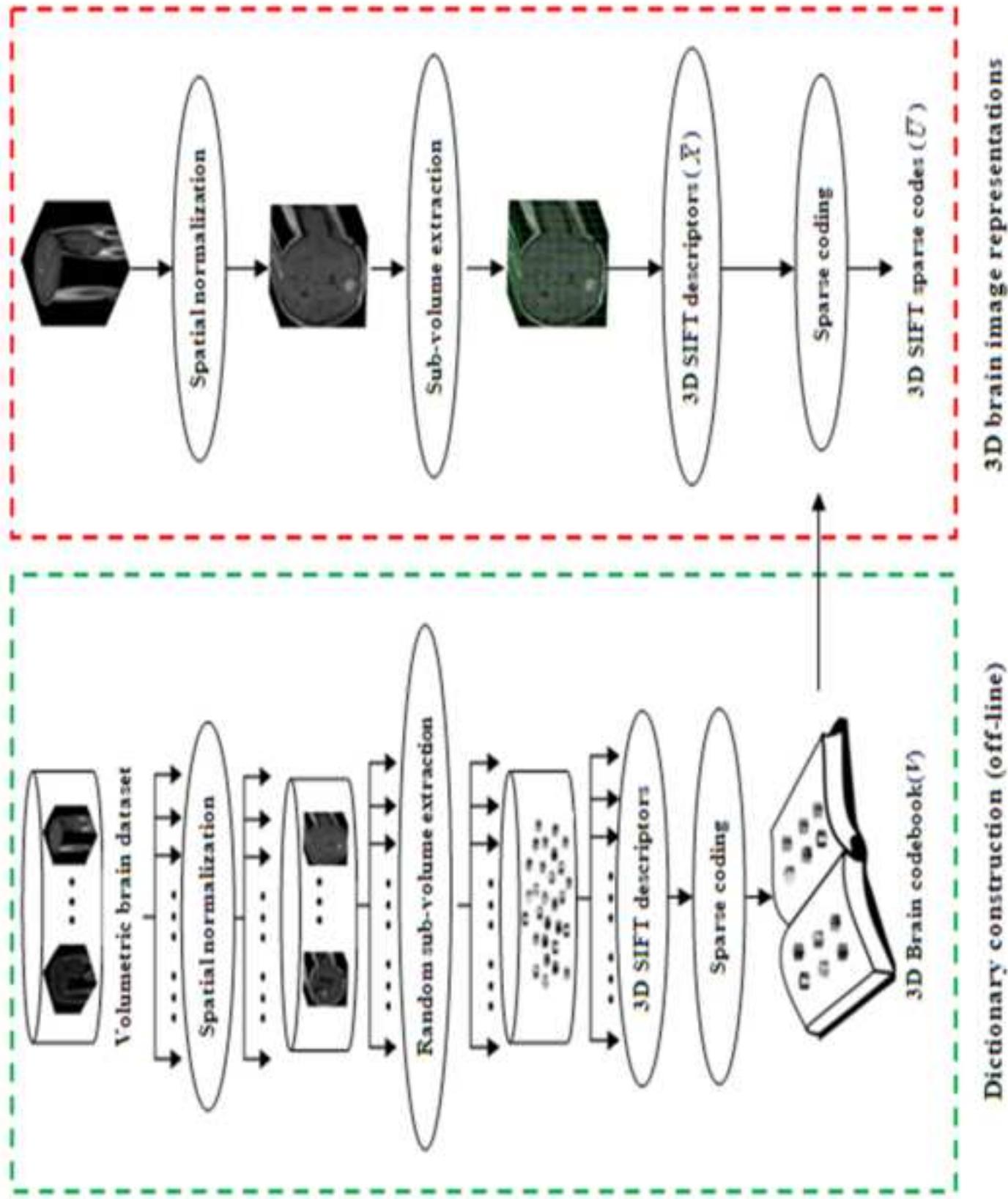


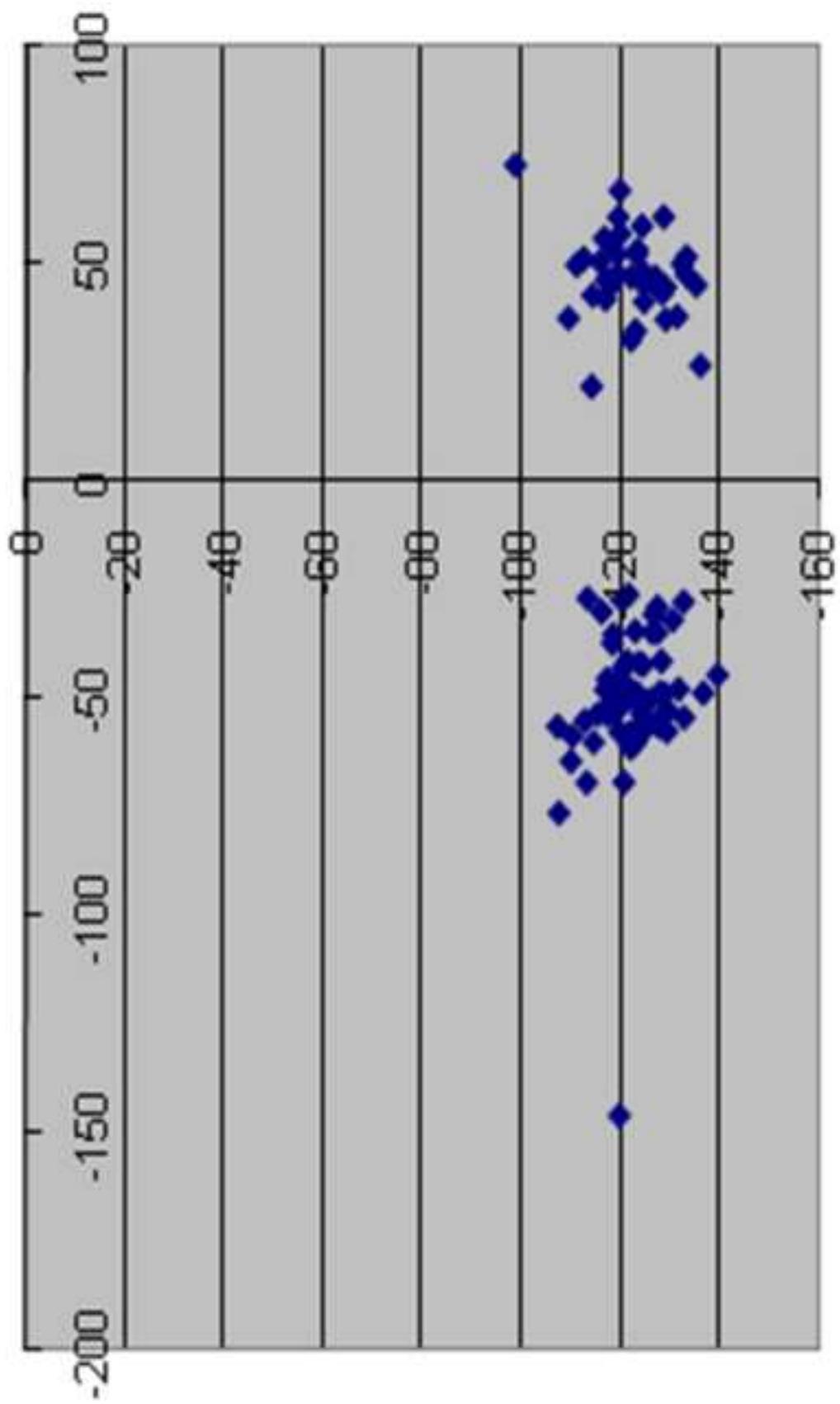
(a) 3D Brain with  $8 \times 8 \times 8$  sub-volumes

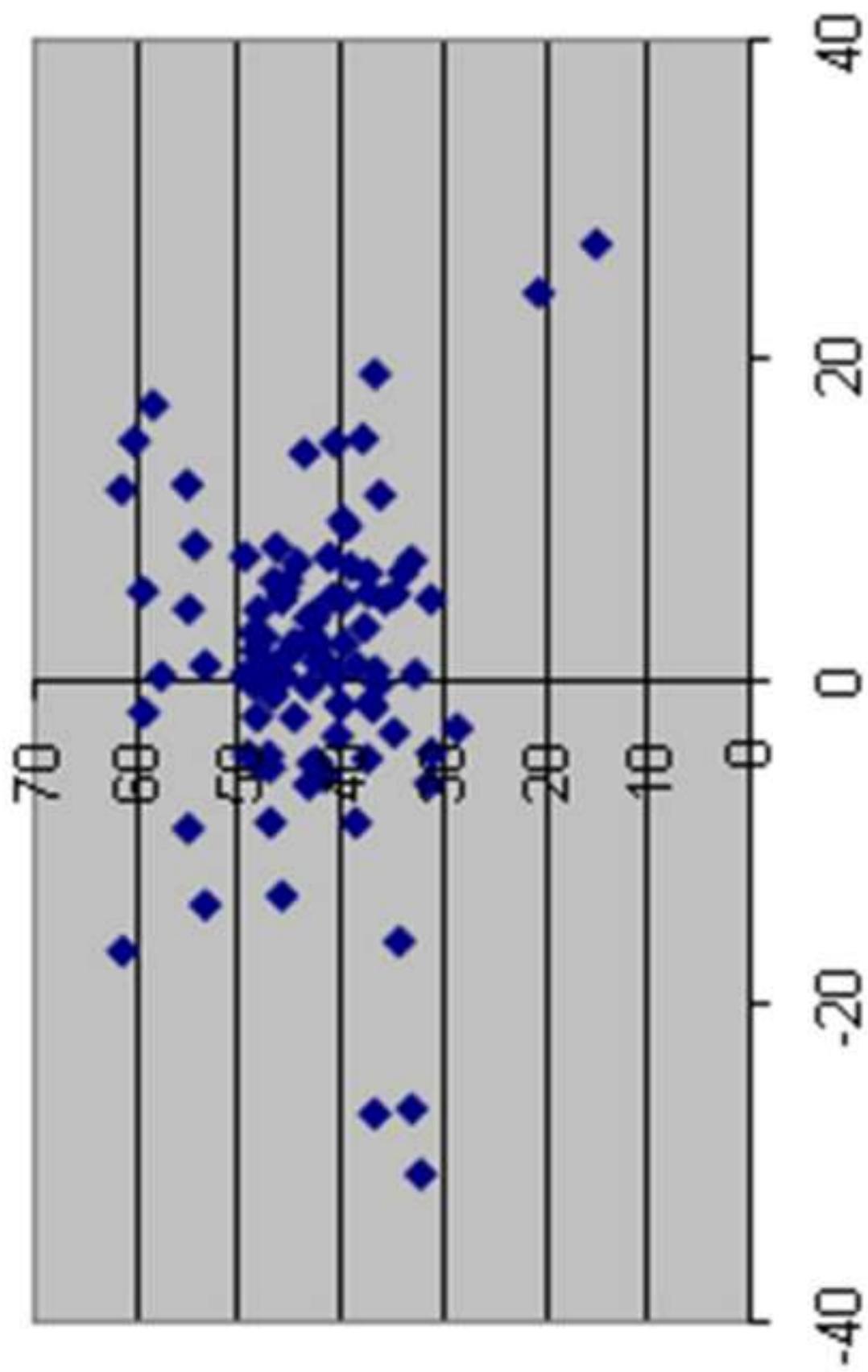
(b) A sub-volume with  $2 \times 2 \times 2$  sub-blocks

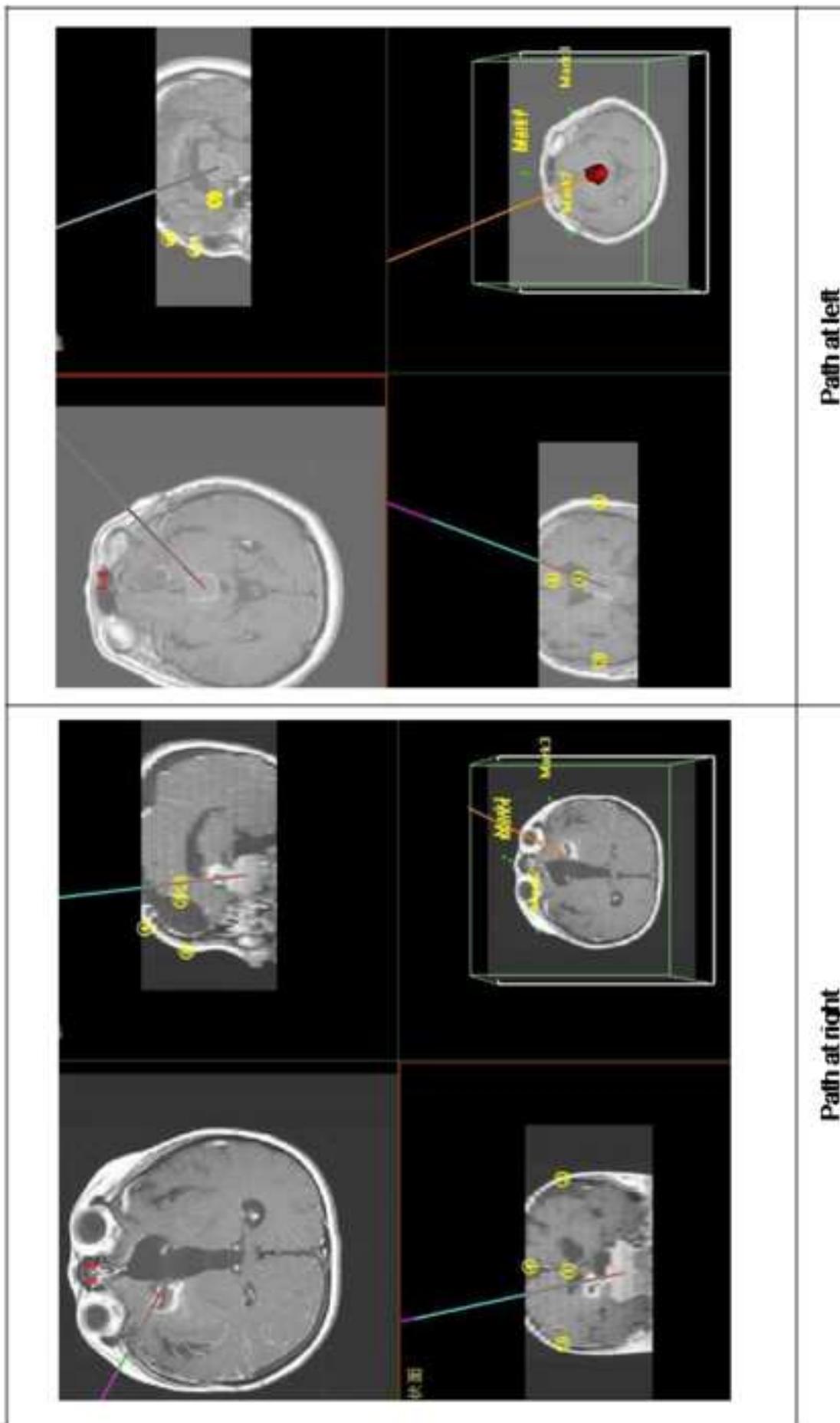
(c) 3D SIFT descriptors ( $X$ ) of a sub-volume

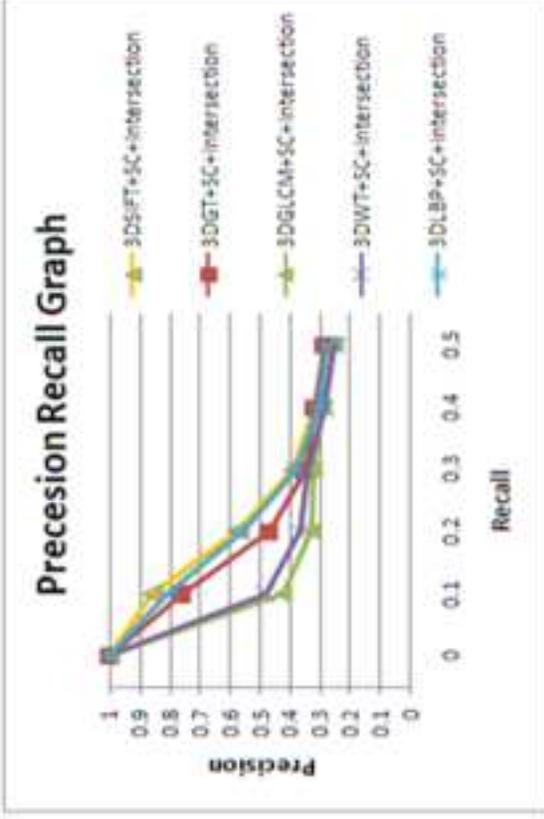
(d) Histogram of gradient orientation



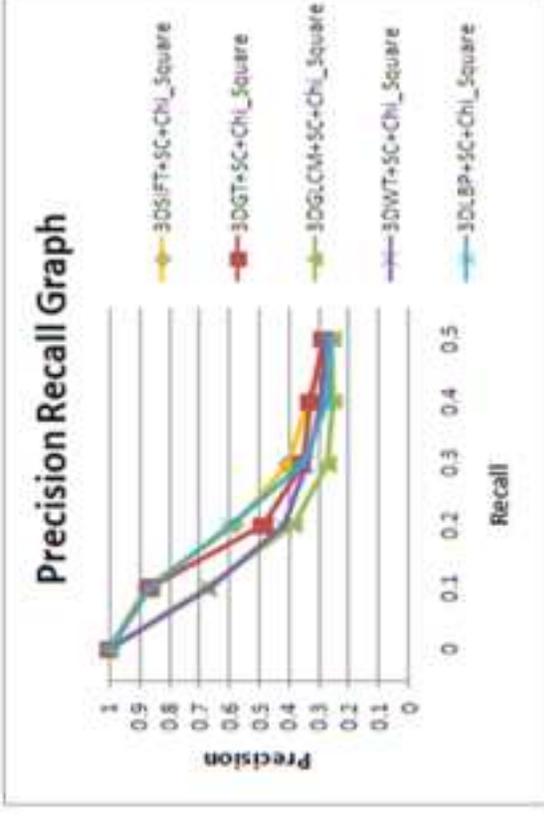




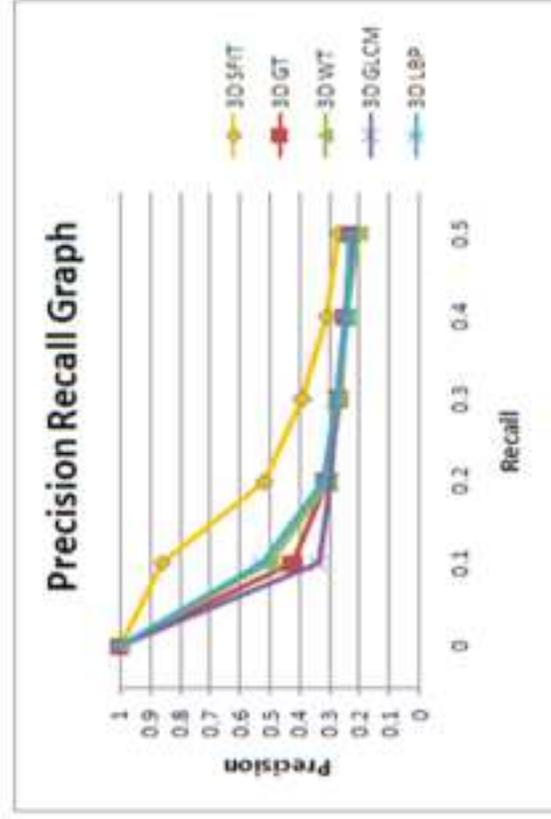




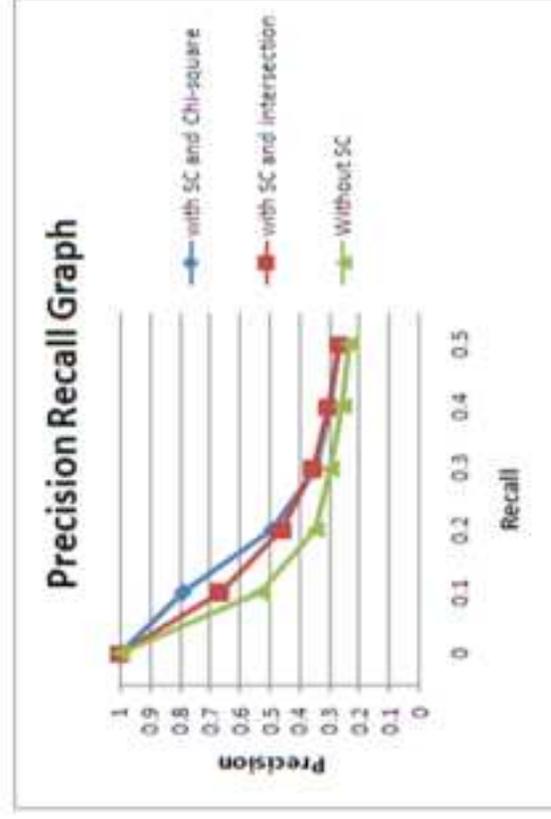
(a) with sparse coding and intersection histogram



(b) with sparse coding and Chi-square histogram



(c) without sparse coding



(d) comparison three methods