TEXTURE-BASED 3D IMAGE RETRIEVAL FOR MEDICAL APPLICATIONS

X. Gao,1 Y. Qian1, R. Hui,2 M. Loomes1, R. Comley1, B. Barn1, A. Chapman3 and J. Rix4

1 School of Engineering and Information Science, Middlesex University, London, NW4 4BT, UK
2 Department of Neurosurgery, General Navy Hospital, Beijing, P.R.China
3 CLQE, Middlesex University, London, NW4 4BT, UK
4 CIE, Middlesex University, London, NW4 4BT, UK

ABSTRACT

Although content-based image retrieval (CBIR) has been researched for more than two decades, retrieving 3D datasets has just begun recently and is mainly focusing on 3D shapes, i.e., iso-surfaces of objects. This is in part due to the fact that the major application area of CBIR is to those images available over the internet. In the medical domain however, more and more images are in three or more dimensions with important information spreading from both shapes to anatomic locations. This study aims to help surgical planning for image-guided neurosurgery, by which the anatomic location plays more important role than the shape of a tumour. Texture-based methods are therefore exploited with four approaches including 3D Grey Level Co-occurrence Matrices (3D GLCM), 3D Wavelet Transform (3D WT), 3D Gabor Transform (3D GT) and 3D Local Binary Pattern (3D LBP) working on a database consisted of around 100 image volumes with both normal and lesioned brains. Preliminary results have shown that 3D LBP performs best with precision recall of 65% and processing time of less than 1 second for both feature extraction and retrieval.

KEYWORDS

3D medical image retrieval, 3D grey level co-occurrence matrices, 3D Wavelet transform, 3D Gabor transform, 3D local binary pattern

1. INTRODUCTION

Due to the advances of medical imaging techniques, more and more images are in three (or higher) dimensional form, giving a coherent and collective view, which however poses a number of challenges when it comes to the management of these images. Since many of these images are consisted of 2D slices, most current databases archive and index them in 2D form, especially for the systems that are indexed by their contents. As a result, a number of limitations have arisen with the significant one being that the information extracted from a single 2D slice can not be representative due to the fact that slices are getting thinner (i.e., resolutions getting higher). Content-based retrieval for three dimensional (3D) images has since been researched widely to meet the demand primarily for the images that are available over the internet, whereas feature descriptions are mainly focusing on 3D shapes, i.e., the iso-surfaces of objects. In this way, invariant of transformation is one of the tasks these methods are targeted at [Novotni2003, Ichida2004, Vajramushti2004, Cao2006, Bustos2007, Gong2009, Mademlis2009]. More recently, 3D object retrieval has been attempted using impact descriptor [Mademlis2009] to capture the surrounding areas of a 3D shape. Because shape-based approach only describes the surface of a 3D object, it ignores the content inside an object. Depth based descriptor is therefore developed as given in [Vajramushti2004] which is however, still capture the outliner of a shape at each depth (z-buffer).

For the application to medical images, volume of interests (VOI) consist of not only boundary shapes, but also inside texture representing tissue properties of the VOI, that equally plays an important role in describing the VOI and should be included in the feature vector as well. Texture-based approaches, on the other hand, including Grey Level Co-occurrence Matrices (GLCM), Wavelet Transform (WT) and Gabor Transform (GT) and Local Binary Pattern (LBP), have been widely used to represent the texture features from images of 2D form, and applied to content-based image retrieval [Müller2004, Unay2008]. In this study,
they are extended to 3D images and are explained in details in the following sections. For 3D volumetric data, i.e., 3D brain images, they comprise of a number of slices “connected together” by the $z$ axis. Although using slice-by-slice 2D approach is still possible, they suffer from the drawback that some important information contained in the volumetric data is missed out. In this research, we explore the aforementioned four 3D texture representations, i.e., 3D GLCM, 3D WT, 3D GT and 3D LBP, to work on 3D brain MR images, and evaluate these methods for 3D medical image retrieval with the application to image-guided neurosurgery.

2. METHODOLOGY

2.1 Volumetric Texture Feature Extraction

2.1.1 3D Grey Level co-occurrence Matrices (3D GLCM)

Gray level co-occurrence matrices are two dimensional matrices of the joint probability of occurrence of a pair of gray values separated by a displacement $d = (dx, dy)$. The displacement can be calculated by a given distance and a direction. In three dimensional form, GLCM can be expanded by specifying a displacement using $d = (dx, dy, dz)$, where $dx$ and $dy$ are the same as described for 2D co-occurrence matrices, and $dz$ representing the number of voxels moved along the $z$-axis of a three-dimensional image [Kovalev2001, Philips2008].

![Figure 1. Thirteen directions in 3D co-occurrence matrices.](image)

In a 3D discrete space, the directions are selected by linking a voxel to each of its nearest 26 (=3*9-1) neighbours respectively, leading to 26 directions. Since each direction and its opposite have the same co-occurrence matrices, we only use 13(26/2=13) directions to describe 3D GLCM, as shown in Figure 1. In this study, four distances in 1, 2, 4, and 8 voxels and thirteen directions are chosen, arriving at 52(=4*13) displacement vectors, and thereafter 52 co-occurrence matrices. As a result, four Haralick texture features [Haralick1973], which are energy, entropy, contrast and homogeneity, are computed from each matrix, giving a feature vector of 3D texture $4$ (measures) * $52$ (matrices) = 208 components, in other words, the dimension of a texture feature vector for a 3D image being 208.

2.1.2 3D Wavelet Transform (3D WT)

Wavelet Transform (WT) provides a spatial and frequency representation of an image, which is similar to the multi-scale way by which the human visual system processes an image. The 2D Wavelet decomposition of an image involves recursive filtering using both high-pass ($H$) and low-pass ($L$) filters along horizontal and vertical directions that is followed by a 2 to 1 sub-sampling of each output image [Mallat1989]. Four Wavelet coefficient images at each scale are generated, i.e., $LL_n$, $LH_n$, $HL_n$ and $HH_n$ subbands respectively. 3D Wavelet transform can be achieved by applying both high-pass (H) and low-pass (L) filters along all three dimensions, leading to eight Wavelet coefficients subbands (one low frequency subband and seven high frequency subbands) at each scale, as shown in Figure 2(a). The process is then repeated in the lowest frequency subband ($LLL_n$), leading to a 3D Wavelet transform of two scales as shown in Figure 2(b).
Finally, the statistical measures (mean $\mu$ and standard deviation $\sigma$) of the Wavelet coefficients in each subband at each scale are computed. We choose 2 scales of 3D Wavelet transforms, as shown in Figure 2. So there are 30 features, i.e., 2(scales) *7 (subbands in each scale) *2 (measures) +2 (measures in the lowest resolution) =30, derived from a Wavelet transform of 2 scales, leading to the dimension of a vector of 3D texture feature being 30.

2.1.3 3D Gabor Transform (3D GT)

Gabor Transform (GT) generates a set of Gabor filters that can be considered as orientation and scale tuneable edge and line (bar) detectors and that therefore are widely applied in texture analysis. A 2D Gabor filter is a complex sinusoid modulated by a Gaussian function whereas a 3D Gabor filter is a direct extension of the 2D form and has been applied in compression of image sequences and face recognition [Wang2005, Feng2007]. The 3D Gabor filter is formulated as Eq. (1).

$$g(x, y, z, F, \theta, \phi) = \hat{g}(x, y, z) \exp \left[ j2\pi(F \sin \theta \cos \phi + F \sin \theta \sin \phi + F \cos \theta) \right]$$

(1)

Where $\hat{g}(x, y, z)$ is a 3D Gaussian function, together with radial centre frequency $F$ and orientation parameters ($\theta$ and $\phi$) determining a Gabor filter in three dimensions. A set of 3D Gabor filters are generated by choosing different $F$, $\theta$ and $\phi$.

In our experiment, we choose four centre frequencies with $F = \{0.0442, 0.0625, 0.0884, 0.125\}$ circle/voxel respectively, six orientation parameters, i.e., $\theta = \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$ and six values of $\phi$, i.e., $\phi = \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$, which gives rise of 144 (= 4*6*6) Gabor filters that are used to extract texture features. Given a 3D volumetric texture $f(x, y, z)$, its 3D Gabor transform $GT_i$ is then defined by

$$GT_i = f(x, y, z) * g(x, y, z, F_i, \theta_i, \phi_i) \quad i = 1, 2, 3 ... 144$$

(2)

The mean $\mu$ and standard deviation $\sigma$ of Gabor transform coefficients are then extracted as a representation of texture features from 144 Gabor transforms respectively. So a feature vector includes 288 components (=4 (scales) *36(orientations) *2 (measures)).

2.1.4 3D Local Binary Pattern (3D LBP)

Local binary pattern (LBP) operator is derived from a general definition of texture in a local neighbourhood. For each pixel in an image, a binary code is produced by thresholding its value with the value of a centred pixel. A histogram is then generated to calculate the occurrences of different binary patterns. In terms of 3D form, Zhao. et al. [Zhao2007] have proposed a 3D dynamic texture recognition by concatenating three histograms obtained from LBP on three orthogonal planes (LBP-TOP), i.e., XY, XZ, and YZ planes, the idea that has been adopted in this investigation.
In this study, the LBP-TOP uses three orthogonal planes that intersect in the centre voxel as shown in Figure 3. We select 8 neighbours with 1 voxel radius as a local neighbourhood. Fifty-nine uniform LBP code is then extracted from the AS, LR and AP planes respectively as shown in Figure 3, producing a 59 bin histogram for each plane by accumulating the 59 binary patterns. Finally, three histograms are concatenated to generate a 3D texture representation, giving the size of a feature vector being 177 (=59*3).

2.2 Retrieval

When a 3D image is submitted as a query, a distance function is used to compare the visual features of two 3D images, \( Q \) and \( I \). For 3D LBP, the histogram intersection is used as a similarity measurement of histogram features that is given in Eq. (3),

\[
D(Q, I) = \sum_i \min(Q_i, I_i)
\]

where \( i \) represents each bin in the histogram. The more similar between query (Q) and I, the bigger value the \( D \) is. Therefore, the retrieved results are ranked in descending order based on the histogram intersection. For the other methods, a normalized Euclidean distance is employed to compare two 3D patterns in a feature space. A normalized Euclidean distance is defined by Eq. (4).

\[
D(Q, I) = \sqrt{\sum_i \frac{(Q_i - I_i)^2}{\sigma_i^2}}
\]

Where \( \sigma_i \) is the standard deviation of a set of representative features over the entire database and are used to normalize the individual feature component. The retrieved 3D images are ranked in ascending order of feature distances.

3. EXPERIMENTS

3.1 Data Collection

In this study, a 3D CBIR based approach is investigated on a collection of 3D brain images (around 100) including both normal and lesioned brains, collected from Neuro-imaging Centre at Beijing General Navy Hospital, China. Each slice has 256 x 256 pixels of cross-sectional gray level with 44 slices from the top of a head to neck, i.e., resolution of 5mm. The images are in DICOM (Digital Imaging and Communications in Medicine) format that has up to 16 bit gray-level resolution. Figure 4 visualizes two 3D brain images (normal and lesion) from the database by using 3DSlicer (http://www.slicer.org/).

Figure 4. Visualization of two 3D brains in database
3.2 Data Pre-processing

In order to describe local features for different parts of a brain, a 3D volume is divided into 8 non-overlapping equally sized blocks, which are 2 blocks along each of x, y, z axes respectively, as shown in Figure 5. Four 3D texture methods described in Section 2.1 are applied to extract local features from each block respectively. Therefore, the dimension of the feature vector for a 3D brain is the size of local features multiplied by 8, leading to 1664, 240, 2304 and 1416 elements in the approached of 3D GLCM, 3D WT, 3D GT and 3D LBP respectively.

Figure 5. Blocks for 3D brain images

4. RESULT

In this section, the performance of 3D medical image retrieval by using four 3D texture methods is evaluated. Figures 6 and 7 illustrate two retrieved results from the above four approaches. The query images are displayed in 3D form and 3 slices appearing in 3 orthogonal planes, i.e., axial, sagittal, and coronal directions. Figure 6 illustrates retrieved results for a query of normal brain, whereas Figure 7 shows the results for a query with a tumour in the middle part. For images with a tumour, the 3D form of representation of each dataset is expressed with the origin located close to the tumour.

Figure 6. Query with normal brain
Figure 7. Retrieved results in top 5 ranking from 3D GLCM (row 1), 3D GT (row 2), 3D LBP (row 3), and 3D WT (row 4).

Figure 8. Query with tumour.

Figure 9. Retrieved results in top 5 ranking from 3D GLCM (row 1), 3D (row 2), 3D LBP (row 3), and 3D WT (row 4).
Figures 10 and 11 demonstrate the average Precision Recall Graph in top ten ranking for ten queries across the whole datasets containing around 100 imagery volumes. Figure 8 illustrates the query with normal brain, while Figure 9 is the query with a tumour in the middle part.

In summary, the mean average precision for twenty queries cross the whole database by using the approaches of 3D GLCM, 3D WT, 3D GT and 3D LBP are 0.6413, 0.5693, 0.6015 and 0.6506 respectively.

Table 1 gives the average processing time of each data volume and query time. The processing time is the average time for extracting the features from a 3D volume. The third column is the average query time for ten queries cross 3D brain database by using four 3D texture methods respectively. All methods run in Matlab R2009a with CPU of Inter P8600 1.58GHz and 3.45GB RAM.

Table 1. Processing and query time

<table>
<thead>
<tr>
<th>Methods</th>
<th>Processing time</th>
<th>Query time</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D GLCM</td>
<td>10.65s</td>
<td>0.83s</td>
</tr>
<tr>
<td>3D WT</td>
<td>2.03s</td>
<td>0.11s</td>
</tr>
<tr>
<td>3D GT</td>
<td>14.3m</td>
<td>0.31s</td>
</tr>
<tr>
<td>3D LBP</td>
<td>0.78s</td>
<td>0.29s</td>
</tr>
</tbody>
</table>

In Table 1, the processing time for 3D GT takes longer time than the other method with 14 minutes, due to the employment of 144 times of 3D convolutions for each block, whereas the processing time for the other methods is only in few seconds.

5. CONCLUSION

The aim of this research is to help surgical planning for image-guided key-hole neurosurgery. Therefore retrieved images with similar locations and sizes of a tumour are more important in the designing of a 3D
CBIR system than shapes, leading to the texture-based investigation of approaches. Four texture based methods, including 3D GLCM, 3D WT, 3D GT and 3D LBP are hence studied. In terms of precision, the four methods perform very similar with recall ranging from 56% to 65%, by which 3D LBP topping the chart having 65% of recall. 3D LBP also carries the fastest processing and retrieval speed with only 0.78 and 0.29 seconds respectively. In Summary, 3D LBP seems to perform the best in this investigation. Due to the fact that we only have a relatively smaller size of database, further study will be in need to verify this finding. Further more, the 3D texture features are very large in terms of the number of elements, the dimension reduction will be considered in the future. In addition, future work will include combinations of some texture descriptors form a better representation to boost retrieval performance.

ACKNOWLEDGEMENT

This research is financially funded by UK JISC. Their support is gratefully acknowledged.

REFERENCES


