

Content-based Retrieval of 3D Medical Images

Y. Qian, X. Gao *, M. Loomes, R. Comley, B. Barn
School of Engineering and Information Sciences
Middlesex University
London, NW4 4BT, United Kingdom

R. Hui, Z. Tian
Department of Neurosurgery,
General Navy Hospital,
Beijing, P.R. China

*Corresponding author: x.gao@mdx.ac.uk

Abstract -- While content-based image retrieval (CBIR) has been researched for more than two decades, retrieving 3D datasets has been progressing considerably more slowly, especially in respect to its application to the medical domain. This is in part due to the limitation of processing speed when trying to retrieve high-resolution datasets in real-time. Another barrier is that most existing methods have been developed based on 2D images instead of 3D, leaving a gap to be filled. At present, a significant number of exploitations are focusing on the extraction of 3D shapes. However, it appears other information tends to be equally important in clinical decision making. In this paper, Local Binary Pattern (LBP), a texture based approach stemming from 2D forms, has been studied extensively through the application to 3D images from a collection of MR brain images in a content-based image retrieval system (CBIR). The initial results show LBP not only can achieve a precision rate of up to 78% but also can perform retrieval in real time with sub-second processing speeds. Comparison with the other three popular texture-based methods, namely 3D Grey Level Co-occurrence Matrices, 3D Wavelet Transforms and 3D Gabor Transforms, is also carried out. The results demonstrate that LBP outperforms them all in terms of retrieval precision and processing speed.

Keywords – CBIR, 3D image retrieval, 3D texture extraction.

1. INTRODUCTION

Due to the advances of medical imaging techniques, more and more images are in three (or higher) dimensional forms, allowing a coherent and collective view. Since many of these images are comprised of 2D slices, most current databases archive and index them in 2D form, especially for the systems that are indexed by their content. As a result, a number of limitations have arisen with the most 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 are getting higher).

On the other hand, at present, content-based retrieval for three dimensional (3D) images has been researched primarily to meet the demand for 3D images over the internet. In this way, the main challenge facing the extraction of features from 3D images is that these features have to be invariant of viewing angles, i.e. invariant of rotation, in order to achieve the retrieval of relevant objects, even though sometimes they may not be visible from all the viewing angles. For example, if a query image is a 3D rabbit with a head facing the view, a good

retrieval system should bring back relevant objects including those showing only its tails as an exact match, i.e. the view angle is at the back of the object. In addition, in 2D cases, the viewing angle is always 0° , being normal to the computer screen, by which most existing algorithms can fulfill this request. The other characteristics of content-based image retrieval (CBIR) are shared between 2D and 3D, including scaling and translation of regions of interest. This has led to many current studies focusing on the invariance of transformations (including rotation, scaling and translation) of objects, which has more to do with shapes. In [1], 3D Zernike descriptors have been developed to describe shapes of objects, by taking advantages of polynomial representations, on which these descriptors are based, being invariant of transformations. In this way, a database has to consist of objects differentiated by shapes, such as airplanes, chairs, etc. Similarly, in order to achieve transformation invariance, a graph-based shape descriptor is created in [2] to determine similarity between 3D objects. More recently, the retrieval of 3D objects has been attempted using impact descriptors [3] to capture the surrounding areas of a 3D shape in order to offer a histogram of time-space curvature that are invariant of rotation and translation. Other shape-based 3D models are included in [4][5][6][7]. Because shape-based approaches only describe the surface of a 3D object, they tend to ignore the content inside that object. Depth based descriptors therefore have been developed as demonstrated in [8], which is however in principle, still capture the outline of a shape at each depth (z-buffer).

For application to medical images, a volume of interest (VOI) consists of not only boundary shapes, but also inside textures representing tissue properties of the VOI. The information extracted from these textures equally plays an important role in describing the VOI and is important to medical doctors most of the time. Therefore these texture features should be taken into consideration in the representation of an object as well.

One way to represent texture is 2D-based, since a 3D dataset is composed of a stack of 2D slices. However, using a slice-by-slice 2D approach suffers from the drawback that some important information inter-laced within the volumetric data is missing. Thus, in terms of a 3D form of texture, this spatial structural information should be extracted from a cube instead of a surface. With this in mind, in this study, the approach of

Local Binary Pattern (LBP) [9] is extended because of its discriminative power and computational simplicity, and applied to a collection of 3D MR brain images for extracting texture information that is subsequently utilized for indexing them. Comparison with the other three popular methods in texture representation is also carried out, including Grey Level Co-occurrence Matrices (GLCM), Wavelet Transforms (WT) and Gabor Transforms (GT). The novelty of this work is the extension of 2D texture features into 3D while minimizing the calculation cost. This is achieved by the introduction of a pre-processing stage of a selection of potential VOIs into query datasets. Through the use of statistically analysis of the bilateral symmetry of a brain MR image, a potential VOI of a query can be detected in real time, before proceeding with the extraction of 3D texture features and the calculation of similarities. This work forms part of our currently online CBIR system at [10]. The structure of the paper is in the following pattern. Section II explains the methods employed in the study, whilst Section III shows the experimental results. The conclusion and discussion are given in Section IV, which is followed by Sections of Acknowledgment and References.

II. METHODOLOGY

In this investigation, at the ingestion of data phase (at least two phases should be in place including ingestion and retrieval from the system), the collected data firstly undergoes a pre-processing stage to normalize them into the same resolution before the indexing stage, as shown in the flow chart in Figure 1. After spatial normalization of volumetric brain data into a standard template, the data are then divided into 64 non-overlapping equally sized blocks, from which, 3D texture features can be extracted to create a feature database. On the query side, a pre-processing stage is introduced to detect a potential VOI after spatial normalization from a query image. Thereafter, 3D texture features from a query are only extracted from these potential sub-blocks of VOIs, which, in the retrieval stage, are compared with the corresponding features in the feature database to obtain retrieval results. Details are explained in the following sub-sections.

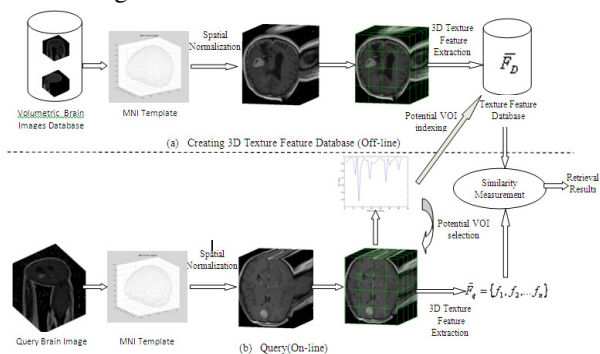


Figure 1. Framework of 3D MR image retrieval.

A. Spatial Normalization

In practice, data are collected from different sources, therefore brain images vary in both shape and size. To make inter-individual brains comparable, it is necessary to transform

the dataset of each individual brain into a standard brain template.

Statistical Parametric Mapping (SPM5) [11] is used in this regard to spatially normalize a brain image to an MNI template [12]. In this way, all the images in the database are of the same size of $157 \times 189 \times 69$ voxels.

B. Extraction of Volumetric Textures

In order to describe local features from different parts of a brain, a 3D volumetric brain is divided into 64 non-overlapping equally sized blocks, giving 4 blocks along each of x , y , z axes respectively, as illustrated in Figure 1. Texture features are then extracted using 3D LBP to create a feature database, upon which image searching and retrieval are performed.

C. 3D Local Binary Pattern (3D LBP)

The Local Binary Pattern operator is derived from a general definition of texture in a local neighbourhood (e.g. 8×8 pixels). In 2D form, for each pixel in an image, a binary code is produced by thresholding its value with the value of a centre pixel. A histogram is then generated to calculate the occurrences of different binary patterns. To extend this method to 3D images, similar to [13], a 3D dynamic texture is recognized by concatenating three histograms obtained from the LBP on three orthogonal planes. When applied to our normalized brain images, they are left-right (LR), Anterior-Posterior (AP), and Superior-Inferior (SI), as shown in Figure 2.

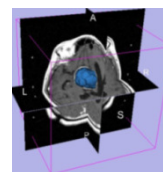


Figure 2. An example of three orthogonal planes in a 3D brain.

These three orthogonal planes intersect in a centre voxel. By selecting 8 neighbours as a local neighbourhood with the radius length being one voxel, fifty-nine uniform LBP codes are subsequently extracted from the planes of SI, LR and AP respectively, again as illustrated in Figure 2, producing a 59 bin histogram for each plane by accumulating 59 binary patterns. Finally, the three histograms are concatenated to generate a 3D texture representation, giving the size of a feature vector as being 177 (59×3) elements.

D. Lesion Detection

The main purpose of the development of this 3D CBIR system is to search images with lesions of similar location, size or shape (all the collections of images are with lesions). Although a feature database has been implemented in advance, the processing of a query has to be conducted in real time, i.e. after a query has been submitted, 3D texture features should be extracted from its 64 sub-volumetric spaces together with the calculation of similarity distances. To this end, while maintaining the overall performance of retrieval, the detection of potential lesions from sub blocks is carried out first to

highlight abnormalities, such as tumours, to speed up the retrieval process.

To do this, the characteristic of bilateral symmetry of a brain along its mid-plane (parallel to SI direction as shown in Figure 2) is assumed. Similar to [14], by comparing the left half with the right half of a hemisphere along this middle symmetry plane, the abnormality is envisaged to be singled out. Since a normalized brain image has been divided into 64 blocks, statistical features (e.g. mean, standard deviation, etc.) of each sub-block together with its mirror block are then calculated and compared to establish potentially abnormal sub-blocks.

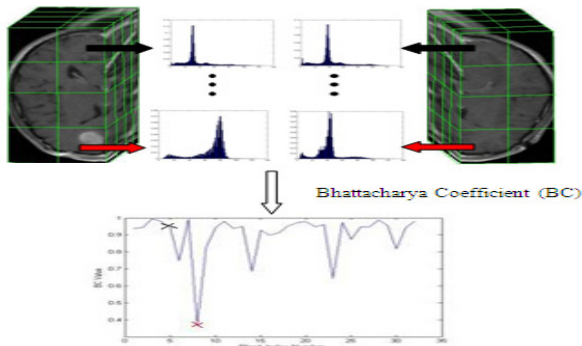


Figure 3. Potential VOI selection

As demonstrated in Figure 3, a normalized brain is divided into left (L) and right (R) parts by a sagittal plane, leading to 32 sub-blocks each, within which a grey level histogram is calculated. The Bhattacharya Coefficient (BC) [15] is then computed between two normalized histogram H_L and H_R , which are obtained from two mirror symmetric sub-blocks as defined in Eq. (1).

$$BC(H_L, H_R) = \sum_i \sqrt{H_L(i) * H_R(i)} \quad (1)$$

The more similar H_L and H_R are, the closer to 1 the BC value is. On the other hand, less similar histograms tend to have smaller BC values. In total, 32 BC values are calculated from 32 paired mirrored symmetric sub-blocks and plotted at the bottom of Figure 3. The horizontal axis points to the index numbers of sub-block pairs whereas the vertical axis represents the corresponding BC values. Also shown in the figure are the BC values presenting the top normal sub-block pair marked with a black 'x', and the bottom abnormal sub-block pair marked with a red cross. Therefore, the mean value of the BC range works as a threshold to be applied to detect the potentially abnormal sub-block (i.e. where $BC < Threshold$).

After the affirmation of a lesioned VOI from a query is established, the 3D texture features are extracted exclusively from this VOI of the query, and are later compared with the features from similar blocks of images in the feature database in an attempt to search images with similar lesions in terms of textures.

E. Similarity Measurement

To measure the degree of similarity between two images Q and I , a distance function should usually be in place calculating the distance between features of two images. For a 3D LBP, the

histogram intersection is applied to measure features of histograms that is given in Eq. (2),

$$D(Q, I) = \sum_i \min(Q_i, I_i) \quad (2)$$

where i represents each bin in the histogram. The more similar they are between a query (Q) and an image I , the bigger the value of the D is. Therefore, the retrieved results are ranked in descending order based on the value of D .

III. EXPERIMENTAL RESULTS

A. Data Collection

The database contains over 100 MR brain images with lesions (e.g. tumour, biopsy) and detailed diagnosis. Each dataset is of resolution of $256 \times 256 \times 44 \text{ mm}^3$, and is in DICOM (Digital Imaging and Communications in Medicine) format with 16 bit grey-level resolution.

B. Results on Detection of Lesions

Since the location of a lesion region plays an important part in retrieving relevant datasets, the evaluation on the detection of lesion positions is carried out first. In Table 1, the first row is the labeling number of the location of a VOI assigned by us for the convenience of calculations, e.g. '1' refers to the abnormal part in the front top left-most part of the brain. The second row is the total number of images containing VOIs in such positions in the database, whilst the number of correctly detected images by the LBP is given on the third row. Therefore the overall performance of the LBP in terms of VOI locations is calculated as the number of detected positive VOIs divided by the total positive VOIs and is 91.3% (168/184).

TABLE 1 VOI DETECTION RATE.

VOI Location	1	2	3	4	5	6	7	8	Total
Number of images	24	46	18	38	24	12	14	8	184
Correctly detected images	24	42	16	34	24	8	12	8	168
Correct Detection Rate (%)									91.3

In terms of precision, the retrieved accuracy is 78% based on ten query images with the ground truth being diagnostic information relating to the locations and sizes of tumours, demonstrating very promising results.

C. Comparison with the Other Texture-based Approaches

The other three methods widely employed in texture representations are also exploited in this investigation by their extension to 3D; including Grey Level Co-occurrence Matrices (GLCM), Wavelet Transforms (WT), Gabor Transforms (GT). These are summarized next.

In 3D form, grey level co-occurrence matrices [16][17] are defined as three dimensional matrices of the joint probability of

occurrence of a pair of grey values separated by a displacement $d = (dx, dy, dz)$.

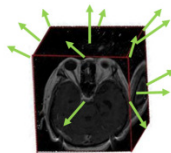


Figure 4. Thirteen directions in 3D GLCM.

For example, four distances with 1, 2, 4, and 8 voxels respectively and thirteen directions, as depicted in Figure 4, and chosen in this study, will produce 52 (4×13) displacement vectors, and thereafter 52 co-occurrence matrices. As a result, four Haralick texture features [18], being energy, entropy, contrast and homogeneity, are computed from each matrix, generating a feature vector with 208 components (4 (measures) × 52 (matrices)).

On the other hand, the 3D WT provides a spatial and frequency representation of a volumetric image, which can be achieved by applying both high-pass (H) and low-pass (L) filters along all three dimensions, which is then followed by a 2 to 1 sub-sampling of each output volumetric image [19], giving rise to eight wavelet coefficients sub-bands (one low frequency sub-band and seven high frequency sub-bands) at each scale, as schematically presented in Figure 5(a). The process is subsequently repeated in the lowest frequency sub-band (LLL_1), providing a 3D wavelet transform of two scales as shown in Figure 5(b).

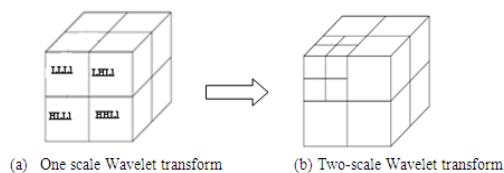


Figure 5. One scale and two scales of 3D WT.

With respect to Gabor Transforms, in order to extend GT into three dimension, a set of 3D Gabor filters are generated similar to [20][21] to detect spatial orientations and scale tunable edges and lines (bar), which can be formulated as Eq. (3).

$$g(x, y, z, F, \theta, \phi) =$$

$$\hat{g}(x, y, z) \exp[j2\pi(F \sin\theta \cos\phi x + F \sin\theta \sin\phi y + F \cos\theta z)] \quad (3)$$

where $\hat{g}(x, y, z)$ is a 3D Gaussian function, together with radial centre frequency F and orientation parameters (θ and ϕ), determining a Gabor filter in three dimensions.

To calculate similarity distances from these three methods, a normalized Euclidean distance is employed to compare two 3D patterns in a feature space, which is defined by Eq. (4).

$$D(Q, I) = \sqrt{\sum_i \left(\frac{Q_i - I_i}{\sigma_i} \right)^2} \quad (4)$$

where σ_i is the standard deviation of a set of representative features over the entire database and are utilized to normalize each individual feature component. The retrieved 3D images are ranked in ascending order of feature distances.

In summary, the above three 3D texture approaches together with LBP are applied to extract texture features from each sub-volumetric block. Furthermore, the dimension of a feature vector for a 3D brain is the size of local features multiplied by 64, the number of blocks each volumetric image is divided into, yielding 13312, 1920, 9216 and 11328 components for the approaches of 3D GLCM, 3D WT, 3D GT and 3D LBP respectively.

The performance of image retrieval is evaluated based on the Precision (P) and Recall (R). Precision is defined as the fraction of retrieved images relevant to a query whilst recall is the fraction of relevant images retrieved. Precision and recall values are usually presented together in a Precision-Recall (P-R) graph, which demonstrates the retrieval performance at each point in the ranking. In a P-R graph, the horizontal axis refers to a recall whereas the vertical axis shows the corresponding precision at each of standard recall points, i.e. 10%, 20%, ..., 100% or 0.1, 0.2, ..., 1. To represent a P-R graph using a single value, usually, the Mean Average Precision (MAP) value is employed to assess the overall performance for all queries and is calculated as:

$$\text{Mean Average Precision (MAP)} = \frac{1}{M} \sum_{i=1}^M AP_i \quad (5)$$

where M is the total number of the queries, AP_i is the average precision for the i^{th} query that is formulated as Eq. (6)

$$\text{AveragePrecision(AP)} = \frac{1}{N_r} \sum_{j=1}^{N_r} P_j \quad (6)$$

where N_r is the total number of relevant images in a dataset for a query, p_j is the precision when retrieving the j^{th} relevant image.

Figures 6 and 7 depict the average Precision Recall Graph for ten queries across the whole datasets with Figure 6 being the results without a pre-processing stage of VOI selection and Figure 7 with the pre-processing stage.

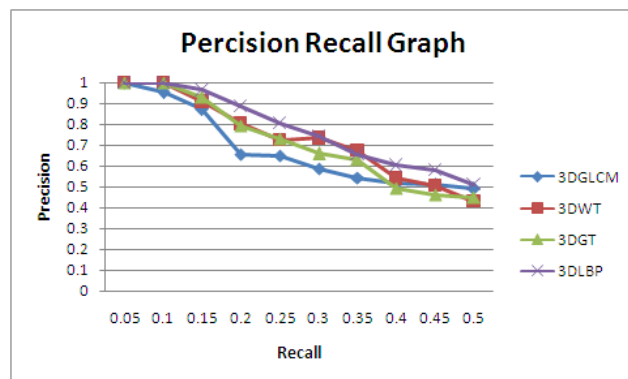


Figure 6. Average precision recall graph for ten queries without VOI selection.

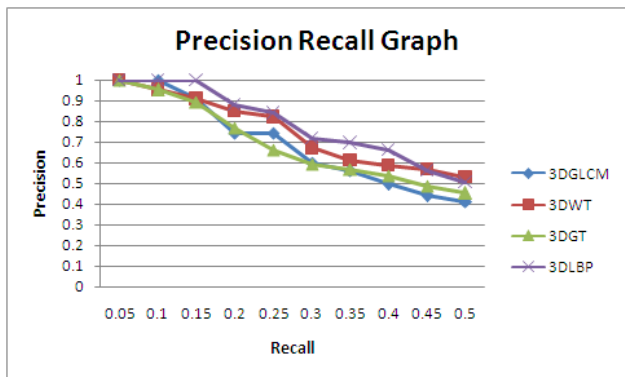


Figure 7. Average precision recall graph for ten queries with VOI selection.

In summary, the mean average precision (MAP) at 0.5 recall rate for ten queries across the whole database by using the approaches of 3D GLCM, 3D WT, 3D GT and 3D LBP are show in the following table.

TABLE 2 VALUE OF MEAN AVERAGE PRECISION

Methods	Without VOI selection	With VOI selection
3D GLCM	0.677	0.690
3D WT	0.731	0.749
3D GT	0.714	0.691
3D LBP	0.774	0.786

Comparing the value of MAP with and without potential VOI selection, the methods of 3D GLCM, 3D WT and 3D LBP with potential VOI selection show a slightly improved performance.

Figures 8 visualizes the retrieved results by using the four approaches with a pre-processing stage of VOI selection. The query image with a tumour in the middle is displayed in 3D fashion and 3 slices appearing in 3 orthogonal planes on the top row, i.e. in axial, sagittal, and coronal directions. The retrieval results are visualized by using an open source software 3D Slicer [3].

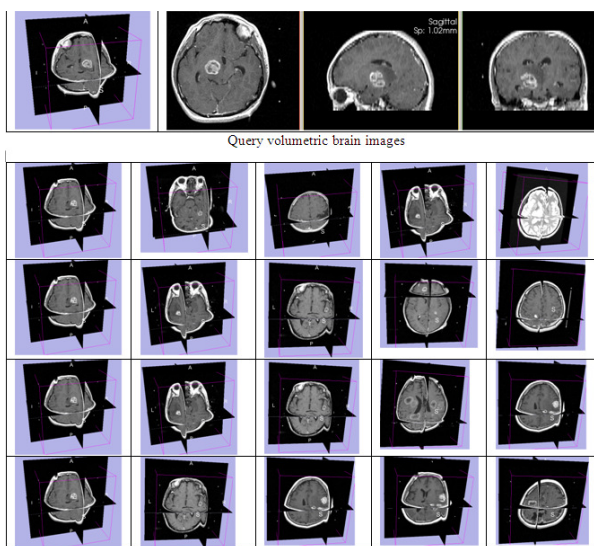


Figure 8. Retrieved results in top 5 ranking from 3D GLCM (row 1), 3D WT(row 2), 3D GT (row 3), and 3D LBP (row 4).

D. Query Time

It is understandable that retrieving images in 3D form might not be performed in real time, one of the drawbacks in the development of CBIR systems for higher dimensions. Table 3 demonstrates the average querying time, amounting to the times spent on both feature extraction and retrieval. The second column is the averaged querying time without a pre-processing stage while the third column is with VOI selection. All methods run in Matlab R2009a on an Intel P8600 1.58GHz CPU with 3.45GByte RAM.

TABLE 3 QUERY TIME

Methods	Without VOI selection	With VOI selection
3D GLCM	43.37s	10.96s
3D WT	4.46s	1.22s
3D GT	38.79m	10.77m
3D LBP	0.74s	0.21s

As can be seen in Table 3, the query time with VOI selection offers 4 times faster operation than that without. In particular, the query time for 3D GT takes a much longer time than the other methods spending 38 minutes, due to the employment of 144 times of 3D convolutions for each block, whereas the query times for the other methods are in the space of few seconds. The table also supports our choice of the 3D LBP approach with sub-second retrieval times and the highest precision rate of 78%, as given in Table 2.

IV. CONCLUSION

In this paper, a texture based approach that draws on the Local Binary Pattern technique has been employed through extension into 3D format, to retrieve lesioned MR brain images in a CBIR system. The results are very encouraging showing that not only higher precision rates can be achieved, but also that it can be done in real time. In comparison with the other three texture based methods, the 3D wavelet approach also performs well with similar retrieval accuracy, but with a poorer query time. In terms of processing speed, it appears the pre-processing stage of detection of potential VOIs is essential to highlight lesions, the regions of interest that retrieved images should contain.

Because of the time required in the establishment of a feature database in 3D form, i.e. normalization, feature extraction, etc., in particular by using the 3D GT approach (up to minutes are needed for each dataset), only ~100 datasets are included in this study. The next step is to process more datasets. Although the precision rate of 78% is very promising, a better rate should be possible with the combination of a few of these texture descriptors, while maintaining the short processing time. Comparison with shape based approaches is also in the pipeline, with the aim of developing CBIR systems for higher dimensional datasets.

ACKNOWLEDGMENT

This research is financially funded by UK JISC. Their support is gratefully acknowledged. The authors would also like

to thank Janet Rix and Alex Chapman at Middlesex University for their contribution to the project.

REFERENCES

- [1] Novotni, M. and Klein, R., "3D Zernike Descriptors for Content Based Shape Retrieval", *Proceedings of the 8th ACM Symposium on Solid Modelling and Applications*, Seattle, Washington, USA, 2003, pp. 216-225.
- [2] Bustos, B. Keim, D., Saupe, D. and Schreck,T., "Content-based 3D Object Retrieval", *In IEEE Transactions on Computer Graphics and Applications*, Vol. 27, No. 4, 2007, pp. 22-27.
- [3] Mademlis, A., Darasb, P., Tzovarab, D., and Strintzis, M.G., "3D Object Retrieval Using the 3D Shape Impact Descriptor", *Journal of Pattern Recognition*, Vol. 42 No.11, 2009, pp. 2447-2459 .
- [4] Cao, L., Liu, J., and Tang, X., "3D Object Retrieval Using 2D Line Drawing and Graph Based Relevance Feedback", *Proceedings of the 14th Annual ACM International Conference on Multimedia*, Santa Barbara, CA, USA, 2006, pp. 105 – 108.
- [5] Ichida, H., Itoh, Y., Kitamura, Y., and Kishino, F., "Interactive Retrieval of 3D Shape Models Using Physical Objects", *Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, NY, USA, 2004, pp. 692 – 699.
- [6] Gong, B., Xu, C., Liu, J. and Tang, X., "Boosting 3D Object Retrieval by Object Flexibility", *Proceedings of the 7th ACM International Conference on Multimedia*, Beijing, China, 2009, pp. 525-528.
- [7] B. Bustos, D. Keim, D. Saupe, Tobias Schreck, Content-Based 3D Object Retrieval, *IEEE Computer graphics and Applications*, 27(4): 22-27, 2007.
- [8] Vajramushti, N., Kakadiaris, I.A., Theoharis, T., and Papaioannou, G., "Efficient 3D Object Retrieval Using Depth Images", *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, New York, NY, USA, 2004, pp. 189 – 196.
- [9] Unay, D., Ekin, A. and Jasinschi, R.S., "Medical Image Search and Retrieval using Local Patterns and Kit Feature Points", *Proceedings of the International Conference on Image Processing*, San Diego, California, USA, 2008, pp. 997-1000.
- [10] <http://image.mdx.ac.uk>.
- [11] <http://www.fil.ion.ucl.ac.uk/spm/>.
- [12] Montreal Neurological Institute, <http://www.mni.mcgill.ca/> .
- [13] Zhao, G. and Pietikainen, M., "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 6, 2007, pp. 915-928.
- [14] Gao, X.W., Batty, S., Clark, J., Fryer, T., Blandford, A., Extraction of Sagittal Symmetry Planes from PET Images, *Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing (VIIP'2001)*, pp 428-433, ACTA Press, 2001.
- [15] Bhattachary A, " On a Measure of Divergence between Two Statistical Populations Defined by Their Probability Distribution", *Bulletin of the Calcutta Mathematical Society*.Vol.35, 1943, pp99-109.
- [16] Kovalev, V.A, Kruggel, F., Gertz, F.J., and Cramon, D. Y., "Three-Dimension Texture Analysis of MRI Brain Datasets", *In IEEE Transactions on Medical Imaging*, Vol. 20, No. 5, 2001,pp. 424-433.
- [17] Philips, C., Li, D., Raicu, D.,and Furst, J., "Directional Invariance of Co-occurrence Matrices within the Liver", *Proceedings of IEEE International Conference on Biocomputation, Bioinformatics, and Biomedical Technologies*, Bucharest, Romania,2008, pp.29-34.
- [18] Haralick, R.M, Shanmugam, K., and Dinstein, I., "Textural Features for Image Classification", *In IEEE Transactions on Systems, Man, and Cybernetics*, Vol.3, No. 6, 1973, pp. 610-621.
- [19] Mallat, S. G., "A Theory for Multiresolution Signal Decomposition: the Wavelet Representation", *In IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No.7, 1989, pp. 674-693.
- [20] Feng, M. and Reed, T.R., "Motion Estimation in the 3-D Gabor Domain", *In IEEE Transactions on Image Processing*, Vol. 16, No. 8, 2007, pp. 2038-2047.
- [21] Wang, Y. and Chua, C., "Face Recognition from 2D and 3D Images Using 3D Gabor Filters", *Journal of Image and Vision Computing*, Vol. 23, No. 11, 2005, pp. 1018-1028.