Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Evaluation of 3D Anatomical Surfaces Indexing for Surgery Planning

Y. Rigollé[1, 2], J. Puentes[1, 3], M. Giordani[2], C. Roux[1, 3]

[1] ENST Bretagne, GET-ENST Département Image et Traitement de l'Information, Brest, France
[2] France Telecom Recherche et Développement, Meylan, France
[3] INSERM, U650, Laboratoire de Traitement de l'Information Médicale, Brest, France

Abstract — **Finding pertinent images in large picture archiving systems, for advanced medical practice support is becoming increasingly difficult. One possible solution to such emerging problem is image indexing. This work proposes to evaluate the indexing and retrieval performance of various 3D anatomical indexing approaches, in order to assist surgery planning based on similar cases. The evaluation examines the indexing performance of 5 feature descriptors (simple statistic, cord-based, shape distribution, surface curvature, and 3D Hough transform) and the retrieval performance of 5 similarity measures (the Minkowski norms L_1, L_2, and L_∞, the Bhattacharyya distance, and the χ^2-divergence). A database of 21 patients, with an average of 11 3D anatomical surfaces per patient was used. The combined performance of feature descriptors and similarity measurements was evaluated with the Bull-Eye Percentage score. Experimental results indicate that there are several possible optimal indexing and retrieval approaches, depending on the surface characteristics.**

Keywords — **3D surface indexing, bull-eye percentage score, surgery planning, anatomical surfaces, MPEG-7.**

I. Introduction

The medical imaging field has been in constant progress, producing affordable devices that provide better-quality images. Besides the growing use of large pictures archiving systems, finding pertinent images in those systems, for advanced medical practice support, is becoming increasingly difficult. One possible solution to this emerging problem is image indexing. Using a similarity measurement, image indexing is intended to classify images of a given image database in a systematic manner. Such approach is gaining relevance in the context of 2D medical images, for computer aided diagnosis [1]. Another field where image indexing is likely to become relevant, is surgery planning. In this context, previously segmented CT scans or MRIs of patient organs, are modeled as 3D surfaces, used to improve surgery's precision and safety through detailed planning and simulation, in addition to their application for virtual surgery training [2]. An example of a planning and surgery simulation tool is Argonaute 3D [3]. It is a medical collaborative application that permits practitioners located in different places to work in real time, on a 3D representation of a patient's organs, in order to establish a diagnostic and define the surgical intervention strategy (Fig. 1).

Fig. 1. Interface example of the Argonaute 3D medical collaborative application.

Integrated to this application, 3D anatomical surfaces indexing could facilitate surgical planning, allowing surgeons to find similar electronic patient records, in the sense of analogous surgical procedures and organ characteristics. To the best of our knowledge, no 3D anatomical surfaces indexing for surgery planning has been carried out until now. Our objective is therefore to identify the most performing surface descriptors and their related similarity measurement, capable of efficiently answer to a similarity request. The proposed evaluation approach only takes into account numerical indexing, which relies exclusively on image values and doesn't handle any textual image description labels, as it is done in semantic indexing. The rest of the paper is organized as follows: section II describes the 3D surfaces database; section III summarizes the performance descriptors and the similarity measures; results of the combined descriptor and similarity measures evaluation are presented in section IV and discussed in section V; section VI states the main findings of our work.

II. 3D Anatomical Surfaces Database

A set of 212 anatomical surfaces, generated after a semi-automatic segmentation of the corresponding anatomical structures CT scans [2], was used to evaluate the performance of the tested anatomical surface indexes and similarity measures. This collection of 3D objects represents 21 different patients, with an average of 11 distinct anatomical surfaces (e.g. heart, lungs, kidneys, or bladder) per patient (Fig. 2).

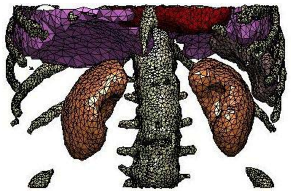

Fig. 2. A patient 3D data set.

The database is consequently structured according to 11 anatomical surfaces classes. Table I lists the different classes along with their name, and indicates the amount of objects that belong to each one.

TABLE I
CATEGORIES OF ANATOMICAL SURFACES

Id	Class name	Size	Id	Class name	Size
1	Aorta	21	7	Left kidney	18
2	Bones	20	8	Right kidney	19
3	Heart	21	9	Right lung	20
4	Liver	21	10	Spleen	20
5	Left lung	21	11	Skin	20
6	Bladder	8		Unclassified	3

III. METHODOLOGY

Our 3D indexing scheme intends to accomplish indexing and retrieval at 2 levels: the first discriminates separated anatomical surfaces according to the previously defined classes, while the second intends to identify each patient anatomical surfaces set with a unique signature, making use of geometrical relations between the different surfaces. This paper focalises on the first one, looking for the optimal approach to index separated anatomical surfaces. In order to index the anatomical surfaces it is necessary to provide (i) a numeric signature that describes the object and (ii) a similarity measure (SM) compatible with the descriptor and capable of evaluating the distance between 2 analogous signatures. There are various types of descriptors in the literature [4]: 3D statistical, extension-based, volume-based, surface geometry-based, image-based, and structural. Given the surface characteristics of the used 3D objects, statistical and surface geometry-based descriptors were identified as the most appropriate for the evaluation. Table II lists the selected feature descriptors (FD).

TABLE II
OVERVIEW OF THE TESTED DESCRIPTORS

Descriptor name	Category
Simple statistic (area, volume)	Statistical
Cord-based descriptor	Statistical
Shape distribution	Statistical
Surface curvature	Surface geometry
3D Hough transform	Surface geometry

Details about the different FDs and their application are presented along with the obtained results in section IV. To find the optimal SM for each descriptor, the performance of the following SMs were compared: Minkowski norms, Bhattacharyya distance and χ^2-divergence.

Minkowski norms, well-known under the name of L_n, are commonly used because of their simplicity and reduced computing time. The evaluated norms L_1, L_2, L_∞ are defined by:

$$L_n(H,G) = \|H-G\|^n = \left(\sum_i |h_i - g_i|^n\right)^{\frac{1}{n}} \quad (1)$$

Where H and G are histogram signatures and $n = 1$, 2 or ∞. The Bhattacharyya distance [5] is calculated as:

$$Bhat(H,G) = 1 - \sum_i \sqrt{h_i} \cdot \sqrt{g_i} \quad (2)$$

Where h_i and g_i are the histogram elements. On the other hand, χ^2-divergence is a statistical method to assess the dissimilarity between 2 probability distributions. In the case of histograms, χ^2-divergence is obtained applying:

$$\chi^2(H,G) = \sum_i \frac{(h_i - g_i)^2}{h_i + g_i} \quad (3)$$

A wide range of criteria can be applied to calculate the {FD, SM} couples accuracy for retrieval purposes [6, 7]: precision-recall graph, distance matrix and scalar scores (first and second-tier, or discounted cumulative gain). Conventionally, precision-recall graphs serve to measure the retrieval quality of indexing methods. *Recall* is the ratio between the number of relevant retrieved records and the total number of relevant records in the database; *precision* is the ratio between the number of retrieved relevant records and the total number of retrieved irrelevant and relevant records. Because of the large number of possible {FD, SM} couples to test, combined with the limitations to visually analyze them, the use of precision-recall graphs is inadequate for our purpose. Fig. 3 illustrates the complexity of the task in the evaluation of 6 cord-based FDs, using the L_1 SM applied to one patient set of anatomical surfaces.

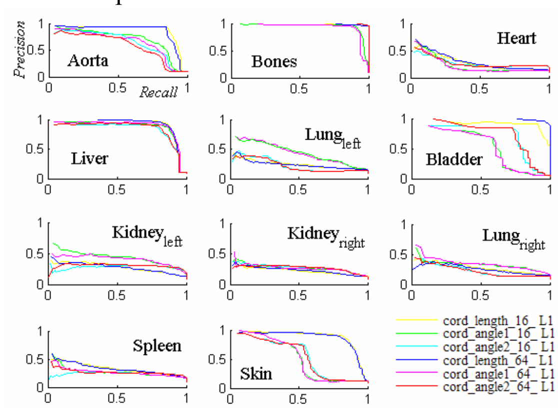

Fig. 3. Recall-precision graphs for 6 cord-based FDs and 1 SM.

Instead, a retrieval quality measurement selected to evaluate the performance of the MPEG-7 standard descriptors, the Bull-Eye Percentage (BEP) score, has been applied. It evaluates the descriptor's performance with a single scalar value. For each query, the BEP is defined as the percentage of elements that belong to the same class, among the $2K$ retrieved results, where K is the size of the class:

$$BEP = \frac{N_{relevant}(2K)}{K} \quad (4)$$

$N_{relevant}(x)$ is the number of relevant records among the x retrieved results.

IV. FD AND SM EVALUATION

This section presents the indexing and retrieval quality results, obtained by computing the BEP scores for each anatomical surface object type, using a FD and the different SMs. Only the value for the SM that outperformed the others in each specific case is displayed, with the resulting BEP score. For each FD, the highest BEP scores indicate the best surface retrieval performances. Those columns are highlighted in each table of the next paragraphs, outlining as well the corresponding FD parameters.

A. Simple statistic

This descriptor deals with the ratio between object volume and area [8]. These 2 values were weighted with varying exponents {*-3*, *3*}. Table III shows the optimal results for the different objects according to the {*k*; *l*} parameters, where *k* is the volume exponent and *l* the area exponent.

TABLE III
VOLUME-AREA RATIO BEP EVALUATION (AVERAGE BEP: 0.667)

Surface	1	2	3	4	5	6	7	8	9	10	11
{k; l}	-1; 1	1; -3	-1; 3	-3; -2	1; -3	1; 2	-1; -1	1; -1	1; -3	-2; 3	-2; -3
SM	L1	L1	L1	L1	L1	L1	L1	L1	L1	L1	L1
BEP	0,721	0,999	0,474	0,905	0,612	0,661	0,435	0,453	0,626	0,447	1,000

B. Cord-based descriptor

This FD requires the utilization of rotation and translation invariance [9]. To obtain the required values a Principal Component Analysis (PCA) is applied to compute the principal object axes, before the FD calculation. Defining a cord as the vector between the centre of mass of an object and a point of its surface, the three versions of the cord-based descriptor have been evaluated. The first version records cord lengths distribution in a histogram, for all points in the mesh ($i = 1..4$ in Table V). The second and the third versions, record the angles distribution between each cord and the first ($i = 5, 6$ in Table V) and second ($i = 7, 8$ in Table V) principal axes, respectively. Table IV shows the 8 types of cord-based descriptors, and table V the obtained results.

TABLE IV
CORD-BASED DESCRIPTORS

Cord dimension	Cord weighting	Normalization	Index i
Length	False	False	1
		True	2
	True	False	3
		True	4
Angle to the first principal axis	N/A	False	5
		True	6
Angle to the second principal axis	N/A	False	7
		True	8

TABLE V
CORD-BASED DESCRIPTOR EVALUATION (AVERAGE BEP: 0,811)

Surface	1	2	3	4	5	6	7	8	9	10	11
{i; j}	3; 64	1; 64	4; 16	5; 64	6; 16	1; 64	5; 64	7; 128	5; 128	1; 64	2; 128
SM	χ^2	χ^2	Bhat	χ^2	Bhat	L2	L∞	χ^2	L1	L1	L1
BEP	0,969	1,000	0,624	0,952	0,707	1,000	0,781	0,611	0,666	0,629	0,982

j is the histogram resolution, which for evaluation purposes takes a value from the set {*16*, *64*, *128*}.

C. Shape distribution

Various features distribution of an object shape, are integrated into the shape descriptors known as *A3*, *D1*, *D2*, *D3*, *D4*. Previous works [10] have determined that the most appropiate descriptor is *D2*. It captures the distribution of distances between random pairs of points on the surface shape. The histogram resolution is also *j*, which for evaluation purposes takes a value from the set {*16*, *64*, *128*}. Table VI lists the obtained results.

TABLE VI
SHAPE DISTRIBUTION DESCRIPTOR EVALUATION (AVERAGE BEP: 0,798)

Surface	1	2	3	4	5	6	7	8	9	10	11
{i ; j}.	2; 128	1; 128	1; 64	1; 128	1; 64	1; 128	1; 128	1; 64	1; 64	1; 128	1; 128
SM	L1	χ^2	L∞	χ^2	L∞	L2	L1	L2	L∞	χ^2	χ^2
BEP	0,967	1,000	0,579	0,952	0,851	0,929	0,539	0,474	0,929	0,553	1,000

D. Surface curvature

This FD deals with the characterisation of the surface local geometric properties. The Koenderink shape index [11], which combines mean and Gaussian curvatures, has been used. On a point *P* of the surface, it is defined as:

$$I(P) = \frac{1}{2} - \frac{1}{\pi}\arctan\left(\frac{k_2(P)}{\sqrt{\left|k_2^2(P) - k_1(P)\right|}}\right) \tag{5}$$

Where k_2 is the mean curvature and k_1 the Gaussian curvature. A histogram *I* is obtained for each surface, with *j* equal to 128. Table VII lists the results.

TABLE VII
SHAPE-BASED DESCRIPTOR EVALUATION (AVERAGE BPE: 0,599)

Surface	1	2	3	4	5	6	7	8	9	10	11
j	128	128	128	128	128	128	128	128	128	128	128
SM	Bhat	L2	Bhat	L1	Bhat	Bhat	Bhat	Bhat	Bhat	L1	Bhat
BEP	0,695	0,787	0,269	0,952	0,438	0,768	0,497	0,427	0,387	0,676	0,698

E. Hough transform

Accumulating the spherical parameters of planes defined by the mesh triangles, the 3D Hough transform generates a 3D histogram, representing the contribution of each triangle, proportional to its area [12]. In the same way as for the cord-based descriptor, 3D objects require a preliminary PCA transform. Moreover, the Hough transform also characterises the local geometric properties of the surface, like the surface curvature descriptor, with an identical signature length *j*, equal to 128. Results are listed in Table VIII.

TABLE VIII
3D HOUGH DESCRIPTOR EVALUATION (AVERAGE BEP: 0,827)

Surface	1	2	3	4	5	6	7	8	9	10	11
j	128	128	128	128	128	128	128	128	128	128	128
SM	L1	L1	L1	L1	Bhat	L1	Bhat	L1	Bhat	L1	Bhat
BEP	0,929	1,000	0,619	0,936	0,810	0,429	0,882	0,795	0,908	0,789	1,000

V. DISCUSSION

Global retrieval performances defined by the different average BEP scores, indicate that the 3D Hough FD has the highest average BEP (82,7%), followed by the cord-based

FD (81,1%), and the D2-descriptor (79,8%). Conversely, the surface curvature descriptor produces a rather poor performance, with a BEP score of 59,9%. Such result is explained by the surface curvature FD high sensitivity to mesh alterations (noise on vertices position, or decimation). In fact, all the 3D objects of the tested database don't have the same mesh quality. For example, the *liver* anatomical surface retrieval, gives an interesting BEP score (95,2%), mainly because those objects were originally created to assist surgeons in the practice of liver tumours surgery [2].

Concerning the retrieval performances related to each anatomical surface, the best BEP results are outlined on tables III, V, VI, VII and VIII. For instance, the best {FD, SM} couple for the aorta is {cord-based, $\chi 2$} with a BEP score of 96,9%. For other surfaces, there are several pertinent {FD, SM} combinations, showing equivalent BEPs, like the objects 2, 4 and 11 (bones, liver, and skin). Assuming a practical application scenario, if prior knowledge (explicit object identification) about the anatomical surfaces is available, the most appropriate {FD, SM} pair could be selected from the previous results. Otherwise, if a limited set of {FD, SM} having an equivalent BEP has been defined, only the simplest computations will be used. Table IX lists the optimal {FD, SM} experimentally identified for each anatomical surface.

TABLE IX
ANATOMICAL SURFACES {FD, SM} CHOICE

	Surfaces	FD	SM	BEP (%)
1	Aorta	Cord	χ^2	96,9
2	Bones	Cord	χ^2	100
3	Heart	Cord	Bhat.	62,4
4	Liver	Cord	χ^2	95,2
5	Left lung	D2	$L\infty$	85,1
6	Bladder	Cord	L2	100
7	Left kidney	Hough	Bhat.	88,2
8	Right kidney	Hough	L1	79,5
9	Right lung	D2	$L\infty$	92,9
10	Spleen	Hough	L1	78,9
11	Skin	Ratio	L1	100

Otherwise, if there isn't prior knowledge about the anatomical surfaces, it could be possible to identify the anatomical surface type from its signature. In this case, table IX indicates for instance, that to test the *heart* surface hypothesis, it is necessary to compute a cord-based signature and to measure its average distances to the indexed database classes with the Bhattacharyya SM. If the mean distance between the *heart* class and the unknown object is the smallest, then the 3D surface is likely to be a *heart*.

Beyond retrieval performance, results suggest that some FDs could be potentially more appropriate to index one type of surface than another. In that sense D2 could be applied for lung description, or the 3D Hough transform for kidney description.

VI. CONCLUSION

This work addresses the retrieval performance of feature descriptors and similarity measures in the context of 3D anatomical surfaces indexing for surgery planning. Potentially optimal {FD, SM} pairs for the indexing and retrieval of different anatomical surfaces, have been evaluated according to the MPEG-7 criterion Bull-Eye Percentage score. Given the significant variety of anatomical surfaces characteristics, there is no single FD or SM capable of efficiently indexing and retrieving all the tested surfaces. However, the obtained preliminary results indicate that combined with prior knowledge, the proposed approach could take advantage of the observed differences, towards the indexation and retrieval of similar patients.

ACKNOWLEDGMENT

The authors would like to thank the IRCAD (Institut de Recherche contre les Cancers de l'Appareil Digestif) in Strasbourg, for providing the 3D images used in this work.

REFERENCES

[1] J.R. Ordonez, G. Cazuguel, J. Puentes, B. Solaiman and C. Roux, "Spatial-textural medical image indexing based on vector quantization," *Proc. 25th Annual International Conference of the IEEE EMBS*, vol. 1, pp. 802-805, 2003.

[2] L. Soler, S. Nicolau, J. Schmid, C. Koehl, J. Marescaux, X. Pennec and N. Ayache, "Virtual reality and augmented reality in digestive surgery," *Proc. third IEEE and ACM International Symposium on Mixed and Augmented Reality ISMAR 2004*, pp. 278-279, 2004.

[3] P. Le Mer, L. Soler, D. Pavy, A. Bernard, J. Moreau, D. Mutter and J. Marescaux, "Argonaute 3D: a real- time cooperative medical planning software on DSL network," *Proc. MMVR12*, pp 203-209, Newport Beach, California, USA, January 2004.

[4] B. Bustos, D. Keim, D. Saupe, T. Schreck and D. Vranic, "An Experimental Effectiveness Comparison of Methods for 3D Similarity Search," *Int. J. Digital Libraries*, vol. 6(1), pp. 39-54, Springer, 2006.

[5] N. A. Thacker, F. J. Aherne and P. I. Rockett, "The Bhattacharyya metric as an absolute similarity measure for frequency coded data," in *Kybernetika*, vol. 32(4), pp. 363-368, 1997.

[6] S. Goodall, P. Lewis and K. Martinez, "3-d shape descriptors and distance metrics for content-based artefact retrieval," *Proc. Storage and Retrieval Methods and Applications for Multimedia*, pp. 87-97, San Jose, California, 2005.

[7] P. Shilane, M. Kazhdan, P. Min and T. Funkhouser, "The princeton shape benchmark," *Proc. Shape Modeling International*, Genova, Italy, pp. 167-178, 2004.

[8] C. Zhang and T. Chen, "Efficient feature extraction for 2d/3d objects in mesh representation," *Proc. International Conference on Image Processing (ICIP)*, vol. 3, pp. 935-938, Thessaloniki, Greece, 2001.

[9] E. Paquet and M. Rioux, "Nefertiti: A Query by Content Software for Three dimensional Models Databases Management," *Proc. First International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97)*, pp. 345-352, 1997.

[10] R. Osada, T. Funkhouser, B. Chazelle and D. Dobkin, "Shape distributions," *ACM Trans. Graphics*, vol. 21(4), pp. 807-832, 2002.

[11] T. Tung and F. Schmitt, "Augmented Reeb Graphs for Content-based Retrieval of 3D Mesh Models," in *International Conference on Shape Modeling and Applications 2004*, pp. 157-166, Genova, Italy, 2004.

[12] T. Zaharia and F. Preteux, "Shape-based retrieval of 3D mesh models," *Proc. 2002 IEEE International Conference on Multimedia and Expo (ICME'2002)*, vol. 1, pp. 437-440, Lausanne, Switzerland, 2002.