MEDICAL IMAGE RETRIEVAL BY CONTENT AND KEYWORD IN A ON-LINE HEALTH-CATALOGUE CONTEXT

[1]*Filip Florea,* [1]*Alexandrina Rogozan,* [1]*Abdelaziz Bensrhair and* [12]*Stefan Darmoni*

filip.florea@insa-rouen.fr

[1] Laboratoire Perception Systèmes et Information, CNRS FRE-2645, INSA de Rouen, France

[2] Équipe CISMeF, Centre Hospitalier Universitaire de Rouen, France

ABSTRACT

In this paper we present keyword and content-based medical image retrieval approaches. Our primary goal is to measure the relevance of our automatic medical image indexing process, which provides us with two signatures: numerical and symbolical. For indexing and retrieval purposes a medical image database, containing six medical modalities (i.e. angiography, ultrasonography, magnetic resonance imaging, standard radiography, computer tomography, and scintigraphy), was created by a medical specialist, from a real healthcare environment. This database was used as a test platform for image feature extraction and modality classification during the indexing stage, and then for performance evaluation during the image retrieval stage. The content-based retrieval showed decent performance, with an average precision of 51% within the 5, 10 or 20 best matches (i.e. most similar to the query image, by a given metric). The modality keyword-based retrieval process yields 88% for both average precision and recall, when an SVM classifier was used to automatically determine the symbolic signatures (i.e. modality class) of medical images. Using the modality information represented by the symbolical signature, allows better adapted image retrieval approaches in a large online medical context.

1. INTRODUCTION

The images attached to health-resources are often of crucial importance in deciding wether a resource is relevant for information retrieval. The CISMeF health-catalogue (French acronym of Catalogue and Index of On-Line Health Resources) provides on-line searching capabilities for health-resources. Adding an image retrieval functionality to the CISMeF health-catalogue will allow the users (i.e. health professionals, students or the general public) to perform better adapted queries depending of what they are searching for.

The CISMeF project [1] was initiated in 1995 in order to meet the users' need to find precisely what they are looking for among the numerous health documents available online[1]. CISMeF describes and indexes the most important resources of institutional health information in French. Indexing is a decisive step for the efficiency of information retrieval within the CISMeF catalogue, and it is also one of the most time consuming tasks for the librarians, demanding high-level documentary skills. Indeed, the textual content of resources is manually annotated with a metadata set and a structured terminology similar to a documentary ontology of the medical field [2]. Ongoing work is concerned with automatic text indexing of medical resources in CISMeF [3].

Being aware of the importance of medical images in healthcare, we currently aim to enrich the health catalogue CISMeF with an image retrieval engine allowing query by keyword and/or by the visual content. Therefore, the cataloguing of medical images extracted from CISMeF resources requires not only numerical-feature extraction (i.e. color, form, texture), but also metadata extraction (i.e. imaging modality, body region or pathology) to form a bimodal: *numerical* and *symbolical signature* of the image visual content.

2. MEDICAL IMAGE RETRIEVAL APPROACHES

Image retrieval has been an extremely active research area over the last 10 years. There are several excellent review articles that are presenting the state-of-the-art of general-purpose CBIR systems like [4][5][6][7][8]. The medical image retrieval systems are generally based on keyword queries and manual textual annotations of the medical images [9], whereas those by visual content are still on a prototype state, dedicated to a very specific medical context and not always accessible via Internet. This makes it impossible to validate and integrate them as effective tools to train or to assist medical students and healthcare professionals in the diagnosis stage.

Most content-based medical image retrieval systems are research prototypes, dealing with a certain modality, a biological system or an anatomical region. Thus, the KMeD [10] and COBRA [11] systems are treating MRI head images, and they are relaying on shape, color, size, texture descriptors and object-based spatial relations extracted from

[1]http://www.cismef.org

regions of interest. ASSERT-system deals with lung CT images [12] by extracting gray-scale, texture and shape descriptors with a human-in-the-loop possibility. I-Browse operates on histological slices [13] using both image and natural language for querying purposes. The system presented in [14] investigates bone X-rays in ophthalmology using shape description and the system presented in [15] describes the retrieval of tumor shapes in mammogram X-rays. The IRMA project is the only one that proposes a general structure for semantic medical image analysis [16] and recently, body-region categorization results have been presented [17]. However, even that it considers multiple modalities, the IRMA system was tested only on x-rays. Given that the principles used by each of these systems are dependent on the particular conditions of diagnosis context, including image modality, they are not directly applicable to other cases.
For this reason we decided to use a bi-modal index that will contain in addition to the numerical signature, a symbolical one, representing the medical modality information. Consequently, knowing the modality will allow us to adapt specific image retrieval methods for each modality in a large medical context (i.e. the CISMeF health-catalogue).
In this paper we present a bimodal image indexing architecture, and performance evaluations for keyword-based and content-based image retrieval on a medical image database extracted from the RUH's (Rouen University Hospital) daily routine.

3. MEDICAL IMAGE DATABASE

The implementation of medical image indexing and retrieval methods requires the constitution of a representative medical image database.
A list of medical image modalities used in daily practice was constituted by a medical expert from the RUH, and implemented in the CISMeF terminology as *resource type* [2]. Since the beginning of this study, the CISMeF team has developed an exhaustive taxonomy of medical image types (N=65) derived from the MeSH tree[2] of diagnosis imaging.
For the experiments presented in this paper, we consider only the main categories of medical-image modalities: standard-angiography, ultra-sonography, magnetic resonance imaging (MRI), standard radiography (X-ray), CT scan (Computer Tomography), and scintigraphy.
The medical images acquired and stored digitally may be very large in size and number. When assembled in image databases, published or even for clinical use, the compression offers a means to reduce the cost of storage and increase speed of transmission. JPEG lossy compression

[2]http://www.nlm.nih.gov/mesh/MBrowser.html

Table 1: Medical Image Database Content

Modality	no. of images	repartition
angiography	337	25.3%
ultra-sonography	180	13.5%
MRI	371	27.9%
X-ray	126	9.5%
CT scan	293	21.9%
scintigraphy	25	1.9%
Total	1332	100%

used almost exclusively on medical image databases involves deliberately discarding information that is not visually or diagnostically important. In this paper our methods are evaluated on a JPEG compressed database, but they can apply to medical images whatever compression scheme is being used.
Our medical image database [Tab. 1] contains 1332 images extracted from Radiology, Radio Pediatry and Nuclear Imaging departments of RUH and of the "Henri Becquerel - Fight against Cancer Center of Rouen". The images do not have the same dimension and quality, being acquired with different digital or analogical equipments, in different hospital services, with different parameters, in a time-period of several years. Thus, the intra-modality variability [Fig. 3], which is usually due anatomical and pathological differences within a modality, is increased. This makes the extraction and selection of an appropriate image feature set even more important, in order to improve the robustness of the automatic indexing process in real-live medical imaging practice.

4. SYSTEM ARCHITECTURE

The medical image retrieval system is based on an automatic indexing stage [Fig. 1]. First, various image feature extraction, feature selection and classification methods have been implemented and compared for medical modality categorization [18]. The best performances have been achieved with an SVM classifier, on a set of texture and statistical features. Prior to classification, the features were selected by considering the consistency within the class (i.e. the modality).
Second, all the medical images extracted from the CISMeF catalogue have to be indexed with two signatures: *one numerical*, representing the selected texture and statistical features used for modality categorization, to describe the visual content of the medical images, and *one symbolical* (the modality categorization result), to indicate the medical modality.
As a result of the medical modality indexing stage, our system allows for image retrieval *by modality keyword* (i.e. "give me all the CT images") and *by content* (i.e.

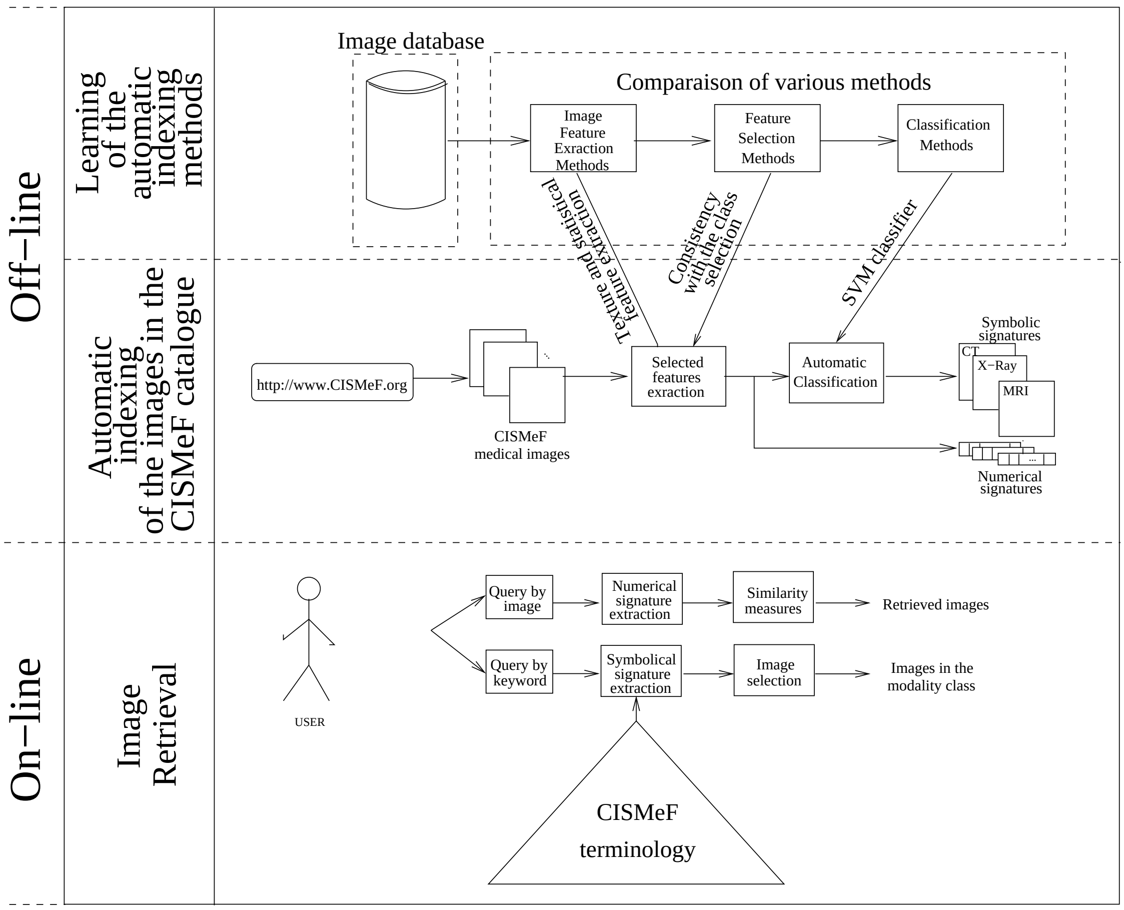

Figure 1: Overview of the system architecture

"give me the images that are the closest to the query-image"). The *keyword* could designate one of the six modalities classes presented in our test-database. The retrieval result will consist of all the images within the modality specified in the query. The CISMeF terminology has to be used to map the natural language queries into the relevant modality keywords. For the *content-based image retrieval* case, similarity measures needed to be computed, between the numerical signature of the initial query-image and those from the database. If i and i_Q are the feature vectors of a given image and of the query image, the similarity between the two is modeled as a distance function. Assuming that the feature vectors are representing accurately the images, the distances between the vectors will measure the dissimilarity between the images. There are series of distance measures developed over the years (the Mahanalobis distance, the Bhattacharyya distance, the Kulback-Leibler divergence, quadratic forms, Chi-square ...) from which we employed the two most frequently used: L_1 (also known as "Manhattan distance") and L_2 (aka "Euclidean distance"), with the following equations:

$$d_{L_1}(i, i_Q) = \sum_{f=1}^{n} |i[f] - i_Q[f]| \quad (1)$$

$$d_{L_2}(i, i_Q) = \sqrt{\sum_{f=1}^{n} (i[f] - i_Q[f])^2} \quad (2)$$

where n is the number of features f.
Once all the distances are extracted, we are usually interested in the first 5, 10 or 20 nearest vectors (i.e. most similar images to the query-image).
If we are interested in retrieving the most similar images to a query image, but only from the images of the same modality, a combination between the retrieval approaches can be used. Given the similarity between certain images of different modalities (i.e. MRI and CT scan, or angiography and X-ray [Fig. 2]), searching only among the images of the same modality as the query image would improve the retrieval result.
In the next section, we will present: the extraction and selection of the image features chosen to represent the *numerical signature* of our image index, and the modality

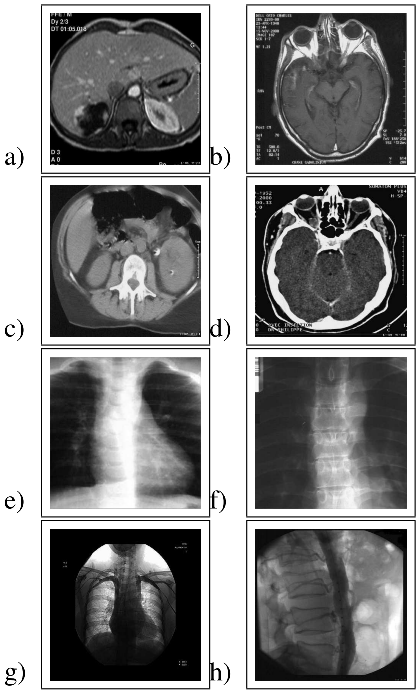

Figure 2: Confusion between a-c, b-d, e-g and f-h a,b=MRI; c,d=CT-scan; e,f=X-Ray; g,h=Angiography

categorization process that will provides us with the *symbolical signature*, representing the modality.

5. AUTOMATIC INDEXING OF MEDICAL IMAGES

5.1. Image feature extraction

The medical indexing architecture we proposed was implemented in a modality categorization context [18]. Due to the important variability [Fig. 3] when taking into account all the medical modalities that could be present into large on-line medical catalogues like CISMeF, there are only few relevant image descriptors that could be used for this purpose. Due to the important variability of anatomical systems and pathologies presented inside a given modality, the color and the shape are either non-invariant (i.e. they do not have invariant characteristics inside a modality) or irrelevant (i.e. they do not help to differentiate the modalities) from the modality point-of-view. Consequently, we have chosen the texture measures and the statistical moments as descriptors for the medical images.
From the large amount of methods developed for describing texture [19], we used *the Harlick's grey-level co-occurrence matrix*, *the fractal dimension* and *the Gabor wavelets*, which seemed to outperform the others. In addition we used features derived from grey-level statistical measures. This way we obtained a 72 feature vector, that contained

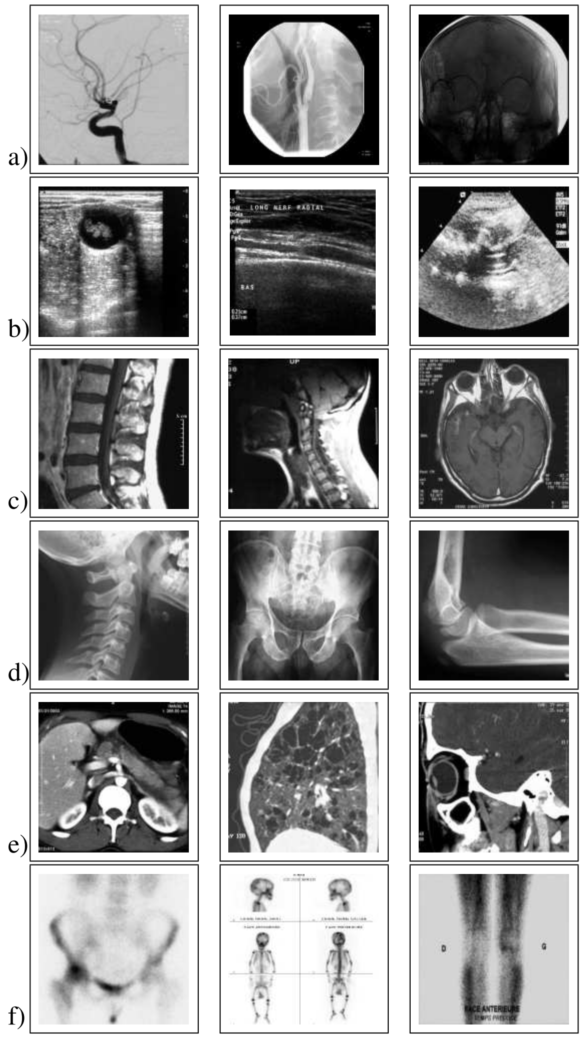

Figure 3: Intra-class variability. a).Angiography; b).Ultrasonography; c).MRI; d).X-Ray; e).CT Scan; f).Scintigraphy

16 coocurrence features (co1-co16), 1 fractal dimension (fd), 48 Gabor features (gb1-gb48) and 7 statistical features (mean, median, mode, L2norm, std, skewness, kurtosis).

5.2. Region of interest segmentation

To avoid the extraction of these features from regions containing non-invariant or irrelevant information between the modalities, a region of interest has to be defined defined. The textual annotations are not always present on medical images, due to legal constraints, but when they are, the text regions have relatively resembling characteristics throughout the modalities. Although the background bears some modality information (i.e. the scintigraphy being mostly presented on a white background), this information is not reliable (i.e. inverted scintigraphy has a black background). Therefore the text and background regions should not be included in the region of interest from where discriminating features are to be extracted.

The textual annotations produced by various medical image systems, were similar enough to be approximated and extracted with a TopHat filter set on the character's thickness [Fig.4b]. This process is enhanced by adding some morphological operations, that consider the horizontal disposition of text in lines, allowing to remove the false detections [Fig.4c]. The background was relatively easy to approximate with the extremities of grey-level histogram [Fig.4d]) and was extracted by thresholding together with some conditions of connectivity to the image borders (i.e. the background regions should have a certain proportion of pixels on the border, with respect to the region's size) [Fig.4e]. Once we obtained the image without the background and text regions, a rectangular 512x512 analysis window centered on the barycenter of the region of interest had been extracted [Fig.4f] and, by our experiments, proved to retain sufficient relevant information for accurate content representation.

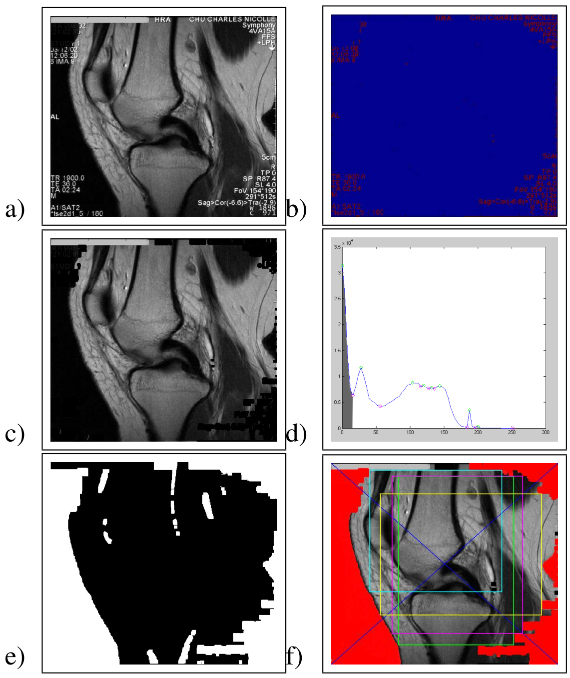

Figure 4: Discarding of the text/background information. a).initial image; b).TopHat filtering; c). removing the text; d).the image grey-level histogram; the background approximation is highlighted in grey; e).the text/background approximation; the errors are corrected with border conditions; f).extraction of the analysis window

5.3. Image feature selection

From all texture and statistical features, the fittest were selected using various feature selection methods, to create a compact and relevant feature set, for a precise and fast classification. We used several feature selection algorithms based on consistency with the class criteria, PCA (principal component analysis), SVM scoring (support vector machines) or entropic-based feature scoring.

The feature selection algorithm that obtained the best performance evaluates (on the full training set) the worth of each subset of features by the level of *consistency* in the class values (i.e. when the training instances are projected onto the subset of features), then searches the space of feature subsets by greedy hill-climbing (augmented with a backtracking facility). By that feature selection algorithm, we reduced the vector size from 72 to 10 features: 4 statistical moments (median, mode, L2norm, kurtosis), 2 cooccurrence (co10, co11), 1 fractal dimension (fd) and 3 Gabor-based features (gb1, gb4, gb25). Furthermore, the feature selection verifies that the four types of features that we choose are complementary to each other: from the initial 72 feature vector, the feature selection method selected 4 out of 7 statistical features, 2/16 co-occurrence features, the fractal dimension and 3/48 Gabor features.

5.4. The modality categorization

As previously stated, our image index contains a symbolical signature representing the image modality. This additional information is important because it will allow us to use better adapted image retrieval (or processing in a larger context) approaches for each modality.

The categorization of medical images according to their corresponding modalities is done by supervised classification of the previously selected features, employing a 10-fold cross-validation scheme. We decided to test the classifiers generalization abilities with a 10-fold cross-validation technique to avoid overfitting. Basically cross-validation means that you use one part of the data (in our case, 9/10) to build a model, which you then apply to the other part (for us, 1/10) of the data to assess how well the model fits the data.

We compared several classifiers (Multi-Layer Perceptron [20], Random Forest [21], Logistic Model Trees [22], Support Vector Machines [23], KNN [24]) and the best results were obtained with an SVM classifier [18]. Several kernel were compared with different penalty coefficients for misclassifications. The best performances were noted with a second degree polynomial kernel, with C=100.

Once the bimodal index is extracted, the performances of the two image retrieval approaches considered, *by keyword* and *by content* can be assessed.

6. IMAGE RETRIEVAL PERFORMANCES

In our present architecture, the keyword-based retrieval permits the extraction of entire modalities using the symbolical signature of the index. Thus, the keyword-based retrieval performance is directly dependent of the modality categorization performance.
The results of the modality categorization process, by SVM classification are given in [Tab. 2]. The confusion matrix reflects the accuracy obtained in our 1332 images, with 10-fold cross validation scheme.

Table 2: Confusion Matrix

a	b	c	d	e	f	← classified as
316	1	4	12	2	2	a = angio
0	177	1	0	2	0	b = ultra-sono
13	3	330	2	23	0	c = MRI
13	2	5	105	1	0	d = X-ray
4	3	29	1	253	3	e = CT-scan
2	0	0	1	2	20	f = scinti

We computed intra-class precision and recall to evaluate image retrieval performances, for the two considered retrieval approaches. Precision is the proportion of relevant retrieved images among all the retrieved images, while recall is the proportion of relevant retrieved images among all relevant images. The first part of the [Tab. 3] summarizes performances of keyword-based retrieval while the second part shows the performances of content-based retrieval within the the 5, 10 and 20 best matches, using the L_1 metrics. The retrieval precisions obtained using L_2 metrics for vector similarity estimation were only 1-2% lower than those obtained with L_1 metrics.
The poorest classification results are obtained for scintigraphy and X-ray classes [Tab. 3] witch are the most under-represented in our database [Tab. 1]. The scintigraphy class has the smallest number of examples, but given the significant visual difference, with regards to the other modalities, the recognition rates are rather high (above 80%) [Tab. 3 - Keyword-based Retrieval]. Therefore, few images are sufficient to learn the scintigraphy class.
Since recall is the proportion of relevant retrieved images among all relevant images, the evaluations of recall among the first 5, 10 or 20 most similar images to the query image, is not interesting.
A retrieval example is presented in [Fig.5]. Considering the two approaches, a knee x-ray content-based retrieval is formulated, on the x-ray class extracted by modality keyword-based retrieval. The query was done with the first image, and the retrieved images are sorted by their L_1 distances to the query image. We observe that among the first 8 retrieved images all are x-rays, and 5 out of 8 are knee-x-rays.

Table 3: Retrieval

Keyword-based Retrieval						
SVM c=100, poly kernel d=2						
	angio	sono	MRI	X-ray	CT	scinti
72 feature vector						
P=	0.908	0.952	0.894	0.868	0.894	0.8
R=	0.938	0.983	0.889	0.833	0.863	0.8
10 feature vector						
P=	0.891	0.917	0.88	0.847	0.879	0.833
R=	0.923	0.983	0.871	0.794	0.843	0.8
Content-based Retrieval						
72 feature vector						
cut	angio	sono	MRI	X-ray	CT	scinti
5 P=	0.550	0.687	0.558	0.457	0.626	0.152
10 P=	0.581	0.741	0.579	0.437	0.680	0.136
20 P=	0.569	0.735	0.563	0.400	0.695	0.118
10 feature vector						
cut	angio	sono	MRI	X-ray	CT	scinti
5 P=	0.429	0.495	0.430	0.250	0.542	0.112
10 P=	0.462	0.521	0.451	0.259	0.589	0.120
20 P=	0.461	0.498	0.433	0.227	0.599	0.092

7. DISCUSSION

We employed only the simplest and well known distance measures, the objective not being the comparison of their performances, but to verify the relevance of the extracted statistical and textural features for medical image retrieval purposes. The content-based retrieval results obtained in L_1 metrics or in L_2 metrics were very similar. In our experiments we obtained better results with query by keyword retrievals that with query by content, due to very accurate symbolic signature extraction. Consequently, the retrieval of entire classes of images by keyword query, is done with high precisions and recalls, nearly 90% of the images being placed in the right class. We observe the similar performances of the two feature vectors: the first composed from all the extracted 72 features and the second composed from the best 10 selected features. The differences in precision and recall between the two are very small, less than 1% being lost by reducing the number of features, all in significantly decreasing the classification time. Concerning the content-based retrieval, we obtain precisions of more than 50% for the most representative modality classes. The reduced performances of the re-

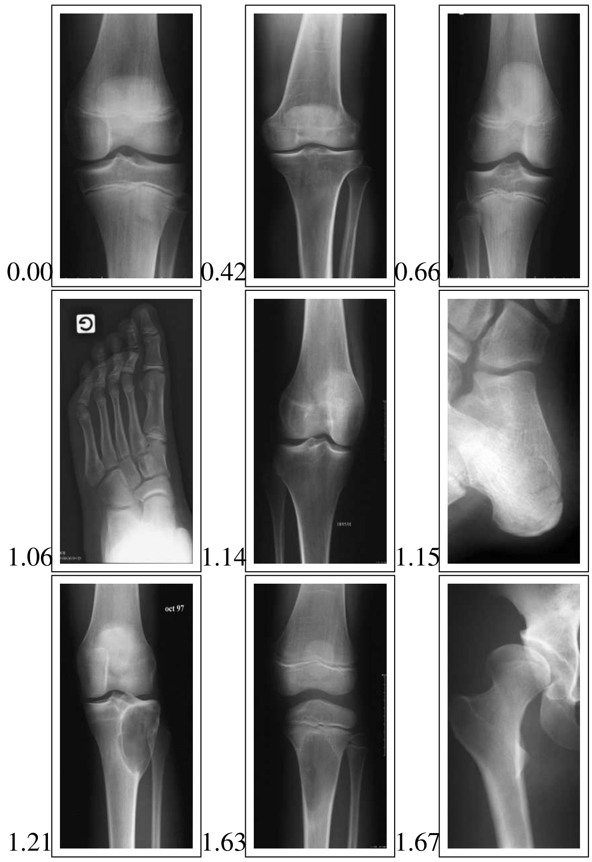

Figure 5: Retrieval example

trieval by content approach can be explained by the fact that the distance computing is not using, unlike SVM, the relative importance of features. The differences between the performances of the two considered vectors rises to approximately 12%, and the difference in computing-time is far less important for distance-evaluation for than for SVM training and classifying.

8. CONCLUSION

In this paper presents evaluations of performances for both keyword and content-based medical image retrieval.

The categorization of the query image, and the optimization of the similarity measure, by weighting, will improve the system performance.

Using a bimodal index, composed from a numerical and a symbolical signature, allows us to perform better adapted image indexing procedures, to treat images from different modalities, anatomical regions or pathologies in a large online medical context.

9. REFERENCES

[1] S.J. Darmoni, J.P. Leroy, B. Thirion, F. Baudic, M. Douyére, and J. Piot, "Cismef: a structured health resource guide," *Meth Inf Med*, vol. 39, no. 1, pp. 30–35, 2000.

[2] M Douyère, L.F. Soualmia, A. Névéol, A. Rogozan, B. Dahamna, J.P. Leroy, B. Thirion, and S.J. Darmoni, "Enhancing the mesh thesaurus to retrieve french online health resources in a quality-controlled gateway," *Health Information and Libraries Journal 2004 (in press)*, 2004.

[3] A. Neveol, A. Rogozan, and S.J. Darmoni, "Automatic indexing of health resources in french with a controlled vocabulary for the cismef catalogue: a preliminary study," *Medinfo*, 2004.

[4] A. Gupta and R. Jain, "Visual information retrieval," *Communications of the ACM*, vol. 40, no. 5, pp. 70–79, 1997.

[5] Y. Rui, T. S. Huang, and S.-F. Chang, "Image retrieval: Past, present and future," in *Proceedings of the International Symposium on Multimedia Information Processing*, M. Liao, Ed., Taipei, Taiwan, 1997.

[6] J. P. Eakins and M. E. Graham, "Content-based image retrieval," Jtap-039, JISC Technology Application Program, Newcastle upon Tyne, 2000.

[7] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.

[8] A. delBimbo, *Visual Information Retrieval*, Academic Press, 1999.

[9] T. Frankewitsch and U. Prokosch, "Navigation in medical internet image databases," *Medical Informatics*, vol. 26, no. 1, pp. 1–15, 2001.

[10] W. Chu, C.C. Hsu, A.F. Cardenas, and R.K. Tiara, "Knowledge-based image retrieval with spatial and temporal constructs," *IEEE Trans KDE*, vol. 10, no. 6, pp. 872–888, 1998.

[11] E. El-Kwae, H. Xu, and M. R. Kabuka, "Content-based retrieval in picture archiving and communication systems," *Journal of Digital Imaging*, vol. 13, no. 2, pp. 70–81, 2000.

[12] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M. Aisen, and L. S. Broderick, "Assert: A physician-in-the-loop content based retrieval system for hrct image databases," *Comp.Vision and Image Understending*, vol. 75, no. 1/2, pp. 111–132, 1999.

[13] H. Y. Tang, Lilian, R. Hanka, H. H. S. Ip, K. K. T. Cheung, and R. Lam, "Semantic query processing and annotation generation for content-based retrieval of histological images," in *International Symposium on Medical Imaging*, San Diego, CA, USA, 2000, vol. 3976 of *SPIE Proceedings*.

[14] W. Zhang, S. Dickinson, S. Sclaroff, J. Feldman, and S. Dunn, "Shape-based indexing in a medical image database," in *Procs IEEE Workshop on Biomedical Image Analysis*, 1998, pp. 221–230.

[15] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas, "Fast and effective retrieval of medical tumor shapes," *IEEE Trans. on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 889–904, 1998.

[16] T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, M. Keysers, D. Kohnen, H. Schubert, and B. B. Wein, "Content-based image retrieval in medical applications for picture archiving and communication systems," in *Medical Imaging*, SPIE Proceedings, Ed., San Diego, California, 2003, 5033, pp. 440–451.

[17] M.O. Güld, D. Keysers, T. Deselaers, M. Leisten, H. Schubert, N. Ney, and T.M. Lehmann, "Comparison of global features for categorization of medical images," in *Proceedings SPIE 2004*, 2004, vol. 5371.

[18] F.I. Florea, A. Rogozan, A Bensrhair, and S.J Darmoni, "Comparison of feature-selection and classification techniques for medical images modality categorization," Tech. Rep. FFI no.01, INSA de ROUEN, Perception Systemes Information FRE CNRS 2645, INSA de Rouen, Septembre 2004.

[19] B. Jähne, H. Haußecker, and P. Geißler, *Handbook of computer vision and applications vol.2*, Academic Press, 1999, Interdisciplinary Center for Scientific Computing University of Heidelberg, Heidelberg, Germany.

[20] H White, "Learning in artificial neural networks: A statistical perspective," *Neural Computation*, vol. 1, no. 4, pp. 425–464, 1989.

[21] Leo Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, October 2001.

[22] N. Landwehr, M. Hall, and E. Frank, "Logistic model trees," *ECML 2003*, 2003.

[23] J. Platt, B. Schölkopf, C. Burges, and A. Smola, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods - Support Vector Learning*. 1998, MIT Press.

[24] D. Aha and D. Kibler, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, 1991.