

IPAL Knowledge-based Medical Image Retrieval in ImageCLEFmed 2006

Caroline Lacoste, Jean-Pierre Chevallet, Joo-Hwee Lim, Xiong Wei, Daniel Raccoceanu,
Diem Le Thi Hoang, Roxana Teodorescu, Nicolas Vuillenemot
IPAL French-Singaporean Joint Lab (I2R, CNRS, NUS, UJF)
viscl, viscjp, joohwee, wxiong, visdaniel@i2r.a-star.edu.sg

Abstract

This paper presents the contribution of IPAL group on the CLEF 2006 medical retrieval task (i.e. ImageCLEFmed). The main idea of our group is to incorporate medical knowledge in the retrieval system within a multimodal fusion framework. For text, this knowledge is in the Unified Medical Language System (UMLS) sources. For images, this knowledge is in semantic features that are learned from examples within structured learning framework. We propose to represent both image and text using UMLS concepts. The use of UMLS concepts allows the system to work at a higher semantic level and to standardize the semantic index of medical data, facilitating the communication between visual and textual indexing and retrieval. The results obtained with UMLS-based approaches show the potential of this conceptual indexing, especially when using a semantic dimension filtering, and the benefit of working within a fusion framework, leading to the best results of ImageCLEFmed 2006. We also test a visual retrieval system based on manual query design and visual task fusion. Even if it provides the best visual results, this purely visual retrieval provides poor results in comparison to the best textual approaches.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing—*Indexing methods, Thesauruses*; H.3.3 Information Search and Retrieval—*Retrieval Models, Information filtering*; H.2 [Database Management]: H.2.4 System—*Multimedia Database*.

General Terms

Measurement, Performance, Experimentation

Keywords

Indexing methods, Thesauruses, Retrieval Models, Information filtering, Multimedia Database.

1 Introduction

Besides the ever-growing amount of medical data produced everyday, medical image retrieval systems have a large potential in medical applications. The three main applications concern medical diagnosis, teaching, and research. For the clinical decision making process, it can be beneficial to find other images of the same modality, of the same anatomic region, and of the same disease [16]. Hence, medical CBIR systems can assist doctors in diagnosis by retrieving images with known pathologies that are similar to a patient's image(s). In teaching and research, visual

retrieval methods could help researchers, lecturers, and student find relevant images from large repositories. Visual features not only allow the retrieval of cases with patients having similar diagnoses but also cases with visual similarity but different diagnoses.

Current CBIR systems [21] generally use primitive features such as color or texture [17, 18], or logical features such as object and their relationships [25, 4] to represent images. Because they do not use medical knowledge, such systems provide poor results in the medical domain. More specifically, the description of an image by low-level or medium-level features is not sufficient to capture the semantic content of a medical image. This loss of information is called the semantic gap. In specialized systems, this semantic gap can be reduced leading to good retrieval results [11, 20, 6]. Indeed, the more a retrieval application is specialized for a limited domain, the smaller the gap can be narrowed by using domain knowledge.

Among the limited research efforts of medical CBIR, classification or clustering driven feature selection and weighting has received much attention as general visual cues often fail to be discriminative enough to deal with more subtle, domain-specific differences and more objective ground truth in the form of disease categories is usually available [8, 15]. In reality, pathology bearing regions tend to be highly localized [8]. Hence, local features such as those extracted from segmented dominant image regions approximated by best fitting ellipses have been proposed [12]. However, it has been recognized that pathology bearing regions cannot be segmented out automatically for many medical domains [20]. Hence it is desirable to have a medical CBIR system that represents images in terms of semantic features, that can be learned from examples (rather than handcrafted with a lot of expert input) and do not rely on robust region segmentation.

The semantic gap can also be reduced by exploiting all sources of information. In particular, mixing text and image information generally increases the retrieval performance [7]. In [2], statistical methods are used for modeling the occurrence of document keywords and visual characteristics. The proposed system is sensitive to the quality of the segmentation of the images. Other initiatives to combine image and text analysis study the use of Latent Semantic Analysis (LSA) techniques [24, 26]. In [24], the author applied the LSA method to features extracted from the two media. The conclusion of this study is that combining the image and the text through the LSA method is not always efficient. The usefulness of LSA is also not conclusive in [26]. Conversely, a simple late fusion of visual and textual indexes provides generally good results.

In this paper, we present our work on medical image retrieval that is mainly based on the incorporation of medical knowledge in the system within a fusion framework. For text, this knowledge is in the Unified Medical Language System (UMLS) sources produced by NLM¹. For images, this knowledge is in semantic features that are learned from examples and do not rely on robust region segmentation. In order to manage large and complex sets of visual entities (i.e., high content diversity) in the medical domain, we developed a structured learning framework that facilitates modular design and extract medical visual semantics. We developed two complementary visual indexing approaches within this framework: a global indexing to access image modality, and a local indexing to access semantic local features. This local indexing does not rely on region segmentation but builds upon patch-based semantic detector [13].

To benefit efficiently from both modalities, we propose to represent both image and text using UMLS concepts in our principal retrieval system. The use of UMLS concepts allows our system to work at a higher semantic level and to standardize the semantic index of medical data, facilitating the communication between visual and textual indexing and retrieval. We propose several fusion approaches and a visual modality filtering is designed to remove visually aberrant images according to the query modality concept(s).

Besides this UMLS-based system, we also investigate the potential of a closed visual retrieval system where all queries are fixed and manually designed (i.e. several examples are manually selected to represent each query).

Textual, visual, and mixed approaches derived from these two systems are evaluated on the medical task of CLEF 2006 (i.e. imageCLEFmed).

¹National Library of Medicine -<http://www.nlm.nih.gov/>

Table 1: Results of textual runs

(a) Automatic runs

Rank	run ID	MAP	R-prec
1/31	IPAL_Textual_CDW	26.46%	30.93%
2/31	IPAL_Textual_CPRF	22.94%	28.43%
3/31	IPAL_Textual_CDF	22.70%	29.04%
5/31	IPAL_Textual_TDF	20.88%	24.05%
10/31	IPAL_Textual_CDE	18.56%	25.03%

(b) relevance feedback run

Rank	run ID	MAP	R-prec
1/1	IPAL_Textual_CRF	25.34%	29.76%

2 UMLS-based Textual Retrieval

UMLS is a good candidate as a knowledge base for medical image and text indexing. It is more than a terminology base because terms are associated with concepts. There exists also different type of links. The base is large (more than 50,000 concepts, 5.5 million of terms in 17 languages), and is maintained by specialists with two updates a year. Unfortunately, UMLS is a merger of different sources (thesaurus, terms lists), and is neither complete, nor consistent. In particular, the links among concepts are not equally distributed. UMLS is a “meta thesaurus”, i.e. a merger of existing thesaurus. It is not an ontology, because there is no formal description of concepts, but its large set of terms and variation restricted to medical domain only, enable us to experiment a full scale conceptual indexing system. In UMLS, all concepts are assigned to at least one semantic type from the Semantic Network. This provides consistent categorization of all concepts in the meta-thesaurus at the relatively general level represented in the Semantic Network. This partially solves the problem of merging existing thesaurus hierarchy during the merging process.

Despite the large set of terms and terms variation available in UMLS, it still cannot cover all possible (potentially infinite) term variation. So we need a concept identification tool that manages terms variation. For English texts, we use MetaMap[1] provided by NLM. We have developed a similar tool developed for French and German documents. This concept extraction tools do not provide any disambiguation. We partially overcome this problem by manually ordering them by thesaurus sources: we prefer source that strongly belong to medicine. For example, this enables the identification of “x-ray” as radiography and not as the physical phenomenon (the wave) which seldom appears in our documents. Concepts extraction is limited to noun phrase (i.e. verbs are not treated).

The extracted concepts are then organized in conceptual vectors, like a conventional vector space IR model. We then use the same weighting scheme provided by our XIOTA indexing system [5].

We tested six retrieval approaches based on this conceptual indexing and an approach - corresponding to run IPAL_Textual_TDF - based on an indexing using MeSH² terms.

Each conceptual text retrieval approach uses a Vector Space Model (VSM) for representing each document and a cosine similarity measure to compare the query index to the database medical report. The *tf · idf* measure is used to weight the concepts. The mapping text-concept is separate for three languages, query vectors of concepts of three languages are fusioned, and used for interrogation using the three indexing separately. Then, the three relevance status values are fusioned together.

One major criticism we have against VSM is the lack of structure of the query. VSM is known

²Medical Subject Headings, MeSH, is the controlled vocabulary thesaurus of the U.S. National Library of Medicine. It is included in the meta-thesaurus UMLS.

to perform well using long textual queries but ignoring query structure. The ImageCLEFmed 2006 queries are rather short. Moreover, it seems obvious to us that it is the *complete* query that should be solved and not only part of it. After query examination, we found out that queries are implicitly structured according to some semantic types (e.g. anatomy, pathology, modality). We call this the "semantic dimensions" of the query. Omitting correct answer to any of these dimensions may lead to incorrect answers. Unfortunately VSM does not provide a way to ensure answers to each dimension.

To solve this problem, we decided to add a semantic dimension filtering step to the VSM, in order to explicitly taking into account the query dimension structure. This extra filtering step retains answers that incorporate at least one dimension. We use semantic structure on concepts provided by UMLS. Semantic dimension of a concept is defined by its UMLS semantic type, grouped into semantic groups: Anatomy, Pathology and Modality. Only a conceptual indexing and a structured meta-thesaurus like UMLS enable us to do such a semantic dimension filtering (DF). This filtering discards noisy answers regarding to the dimension query semantic structure. The corresponding run is "IPAL_Textual_CDF". We also test a similar dimension filtering based MESH terms (run "IPAL_Textual_TDF"). In this case, the association between MeSH terms and a dimension had to be done manually. According to Table 1, using UMLS concepts more than terms improves the results of 2 Mean Average Precision (MAP) points (i.e. from 21% to 23%).

Another solution to take into account query semantic structure is to re-weight answers according to dimensions (DW). Here, Relevance Status Value output from VSM is multiplied by the number of concepts matched with the query according to the dimensions. This simple re-weighting scheme strongly emphasizes the presence of maximum number of concepts related to semantic dimensions. This re-weighting step implicitly do the previous dimension filtering (DF), as the relevance value is multiplied by 0. According to our results in Table 1 this DW approach - corresponding to the run "IPAL_Textual_CDW" - produces the best results of 2006 ImageCLEFmed with 26% of MAP. This result outperforms any other classical textual indexing reported in ImageCLEFmed 2006. Hence we have shown here the potential of conceptual indexing.

In run "IPAL_Textual_CPRF", we tested Pseudo-Relevance Feedback (PRF). From the result of late fusion of text-image retrieval results, three top relevant documents retrieved are taken and all concepts of these documents are added into query for query expansion. Then, a dimension filtering is applied. In fact, this run should have been classified in the mixed runs as we also use the image information to have a better precision in the three first images. This PRF approach improves slightly the results obtained with a simple dimension filtering. This is principally due to the fact errors can be present in the three first documents, even with the best mixed retrieval result. Using a manual Relevance Feedback (RF), we obtained a MAP of 25% that is 2 points higher than the result obtained with a simple dimension filtering. In this last run - named "IPAL_Textual_CRF" - a maximum of 4 top relevant documents were chosen by human judgment over 20 first retrieved image. All the concepts from these documents are added into the query for query expansion.

We also tested document expansion using the UMLS semantic network. Based on UMLS hierarchical relationships, each database concept is expanded by concepts positioned at a higher level in the UMLS hierarchy and connected to this concept with respect to the semantic relation "is a". The expanded concepts have a higher position than document concept in UMLS hierarchy. For example a document indexed by the concept "molar teeth" would be also indexed by the more general concept "teeth". This document would be thus retrieved if the user ask for a teeth photography. This expansion does not seem relevant according to the Table 1 as the run "IPAL_Textual_CDE" - that uses a document expansion technique and a dimension filtering - is 4 points below the simple dimension filtering.

3 Visual Retrieval

3.1 UMLS-based visual indexing and retrieval

In order to manage large and complex sets of visual entities in the medical domain, we developed a structured learning framework to facilitate modular design and learning of medical semantics from images. This framework allows to index images using VisMed terms, that are typical semantic tokens characterized by a visual appearance in medical image regions. Each VisMed term is expressed in the medical domain as a combination of UMLS concepts. In this way, we have a common language to index both image and text, which facilitates the communication between visual and textual indexing and retrieval. We developed two complementary indexing approaches within this statistical learning framework:

- a global indexing to access image modality (chest X-ray, gross photography of an organ, microscopy, etc.);
- a local indexing to access semantic local features that are related to modality, anatomy, and pathology concepts.

After a presentation of both approaches in Sections 3.1.1 and 3.1.2, retrieval procedures and experimental results are given in Section 3.1.3.

3.1.1 Global UMLS Indexing

The global UMLS indexing is based on a two level hierarchical classifier according to mainly modality concepts. This modality classifier is learned from about 4000 images separated in 32 classes: 22 grey level modalities, and 10 color modalities. Each indexing term is characterized by a UMLS modality concept (e.g. chest X-ray, gross photography of an organ) and, sometimes, a spatial concept (e.g. axial, frontal, etc), or a color percept (color, grey). The training images come from the CLEF database (about 2500 examples), from the IRMA³ database (about 300 examples), and from the web (about 1200 examples). The training images from ImageCLEFmed database was obtained from modality concept extraction using medical reports. A manual filtering step on this extraction process to remove irrelevant examples had to be performed. We plan to automate this filtering in the near future.

The first level of the classifier corresponds to a classification for grey level versus color images. Indeed, some ambiguity can appear due to the presence of colored images, or the slightly blue or green appearance of X-ray images. This first classifier uses the first three moments in the HSV color space computed on the entire image. The second level corresponds to the classification of modality UMLS concepts given that the image is in the grey or the color cluster. For the grey level cluster, we use grey level histogram (32 bins), texture features (mean and variance of Gabor coefficients for 5 scales and 6 orientations), and thumbnails (grey values of 16x16 resized image). For the color cluster, we have adopted HSV histogram (125 bins), Gabor texture features, and thumbnails. Zero-mean normalization [9] was applied to each feature. For each SVM classifier, we adopted a RBF kernel:

$$\exp(-\gamma|x - y|^2) \quad (1)$$

where $\gamma = \frac{1}{2\sigma^2}$ and with a modified city-block distance:

$$|x - y| = \frac{1}{F} \sum_{f=1}^F \frac{|x_f - y_f|}{N_f} \quad (2)$$

where $x = \{x_1, \dots, x_F\}$ and $y = \{y_1, \dots, y_F\}$ are feature vectors, x_f, y_f are feature vectors of type f , N_f is the feature vector dimension, and F is the number of feature types: $F = 1$ for the grey versus color classifier, $F = 3$ for the conditional modality classifiers: color, texture, thumbnails.

³http://phobos.imib.rwth-aachen.de/irma/index_en.php

We use $\gamma = 1$ in all our experiments. This just-in-time feature fusion within the kernel combines the contribution of color, texture, and spatial features equally [14].

The probability of a modality MOD_i for an image z is given by:

$$P(\text{MOD}_i|z) = \begin{cases} P(\text{MOD}_i|z, C)P(C|z) & \text{if } \text{MOD}_i \in C \\ P(\text{MOD}_i|z, G)P(G|z) & \text{if } \text{MOD}_i \in G \end{cases} \quad (3)$$

where C and G denote the color and the grey level clusters respectively, and the conditional probability $P(\text{MOD}_i|z, V)$ is given by:

$$P(c|z, V) = \frac{\exp^{\mathcal{D}_c(z)}}{\sum_{j \in V} \exp^{\mathcal{D}_j(z)}} \quad (4)$$

where \mathcal{D}_c is the signed distance to the SVM hyperplane that separate class c from the other classes of the cluster V .

After learning - using SVM-Light software⁴ [10, 22] -, each database image z is indexed according to modality given its low-level features z_f . The indexes are the probability values given by Equation (3).

3.1.2 Local UMLS Indexing

To better capture the medical image content, we propose to extend the global modeling and classification with local patch classification of local visual and semantic tokens (LVM terms). Each LVM indexing term is expressed as a combination of Unified Medical Language System (UMLS) concepts from Modality, Anatomy, and Pathology semantic types. A Semantic Patch Classifier was designed to classify a patch according to the 64 LVM terms. In these experiments, we have adopted color and texture features from patches (i. e. small image blocks) and a classifier based on SVMs and the softmax function [3] given by Equation (4). The color features are the three first moments of the Hue, the Saturation, and the Value of the patch. The texture features are the mean and variance of Gabor coefficients using 5 scales and 6 orientations. Zero-mean normalization [9] is applied to both the color and texture features. We adopted a RBF kernel with modified city-block distance given by Equation (2). The training dataset is composed of 3631 patches extracted from 1033 images mostly coming from the web (921 images coming from the web and 112 images from the ImageCLEFmed collection $\sim 0.2\%$).

After learning, the LVM indexing terms are detected during image indexing from image patches without region segmentation to form semantic local histograms. Essentially, an image is tessellated into overlapping image blocks of size 40x40 pixels after size standardization. Each patch is then classified into one of the 64 LVM terms using the Semantic Patch Classifier. An image containing P overlapping patches is then characterized by the set of P LVM histograms and their respective location in the image. An histogram aggregation per block gives the final image index : $M \times N$ LVM histograms. Each bin of the histogram of a given block B corresponds to the probability of a LVM term presence in this block. This probability is computed as follows:

$$P(\text{VMT}_i|B) = \frac{\sum_z |z \cap B| P(\text{VMT}_i|z)}{\sum_z |z \cap B|} \quad (5)$$

where B is a block of a given image, z denotes a path of the same image, $|z \cap B|$ is the area of the intersection between z and B , and $P(\text{VMT}_i|z)$ is given by Equation (4). To facilitate spatial aggregation and matching of image with different aspect ratios ρ , we design 5 tiling templates, namely $M \times N = 3 \times 1, 3 \times 2, 3 \times 3, 2 \times 3$, and 1×3 grids resulting in 3, 6, 9, 6, and 3 probability vectors per image respectively.

⁴<http://svmlight.joachims.org/>

3.1.3 Visual retrieval using UMLS-based visual indexing

We propose three retrieval methods from query by example(s) based on the two UMLS-based visual indexing. When several images are given in the query, the similarity between a database image z with the query is given by the maximum value among the similarities between z and each query image.

The first method - corresponding to run “IPAL_Visual_MC” - is based on the global indexing scheme according the modality. An image is represented by a semantic histogram, each bin corresponding to a modality probability. The distance between two images is given by the Manhattan distance (i.e. city-block distance) between the two semantic histograms.

The second method - corresponding to run “IPAL_Visual_SPC” - is based on the local UMLS visual indexing. An image is then represented by $M \times N$ semantic histograms. Given two images represented as different grid patterns, we propose a flexible tiling (FlexiTile) matching scheme to cover all possible matches [13]. The distance between a query image and a database image is then the mean of block by block distances on all the possible matches. The distance between two blocks is given by the Manhattan distance between the two LocVisMed histograms.

The last visual retrieval method - corresponding to run “IPAL_Visual_SPC+MC” - is the fusion of the two first approaches. This approach combines thus two complementary sources of information, the first concerning the general aspect of the image (global indexing according to modality), the second concerning semantic local features with spatial information (local UMLS indexing). The similarity to a query is given by the mean of the similarity to a query according each index.

The 2006 CLEF medical task was particularly difficult this year for purely visual approaches. Indeed, the queries were at a high semantic level for a general retrieval system. As a proof, the best automatic visual result was less than 8% of MAP. Mixing the local and global indexing, gives us the third place with 6% of MAP as showed in Table 2⁵. We believe than we can improve these results using also the textual query in the retrieval process. Indeed, besides the usual similarity-based queries, our semantic indexing allow semantic-based query. Tests are in course on 2005 and 2006 medical tasks, providing promising results.

Table 2: Results of automatic visual runs

Rank	run ID	MAP	R-prec
1/11	CINDL_Fusion_Visual	07.53%	13.11%
3/11	IPAL_Visual_SPC+MC	06.41%	10.69%
4/11	IPAL_Visual_MC	05.66%	09.12%
6/11	IPAL_Visual_SPC	04.84%	08.47%

3.2 Manual Query Construction and Visual Task Fusion

To see how far we can go with a purely visual approach, we propose here a closed visual system based on manual query construction and visual task fusion. This work is similar to what we did in ImageCLEFmed 2005[23]. We fused retrieval results generated by systems using multiple-feature representations and multiple retrieval systems. More specifically, we used three types of feature representations, i.e., “blob”, “icon” and “blob+icon” and two retrieval systems, “SVM” and “Dist”. For each topic, we manually chose about 50 similar images which is used to form a training set for SVM and to construct the query. All images are then represented by these features respectively and passing either retrieval system.

⁵The Mean Average Precision and the Recall-precision computed in imageCLEFmed for the run IPAL_Visual_SPC+MC was 6.34% and 10.48% respectively because we only submitted - by error - the 25 first queries.

In this year’s attempts, we have submitted 10 runs based on the fusion of six sub-runs. The generated sub-runs are denoted D1,D2,D3,D4, D5 and D6. D1,D2 and D3 use “Dist” retrieval system but different features (D1 using “icon”, D2 using “blob”, D3 using “blob+icon”), D4 and D5 use “SVM” with different features (D4 using “blob”, D2 using “icon”), and D6 uses the UMLS-based system presented in Section 3.1 (D6 corresponds to the run “IPAL_Visual_MC” that is based on global UMLS indexing).

Different from the work done in 2005, we also use the probability estimation of each image about its modality that is given by Equation (3). Some of these sub-runs (D1-D6) are linearly combined together to produce a score for each image. Each score may then be multiplied by the probability and all the results are sorted to yield the final retrieval ranking lists. The runs “IPAL_CMP_D1D2D4D5D6”, “IPAL_CMP_D1D2D3D4D5”, “IPAL_CMP_D1D2D3D4D5D6”⁶, and “IPAL_CMP_D1D2D4D5” used the probability estimations. We also applied a color filter to remove those images whose number of color channels are less than that of the query images, except for runs “IPAL_D1D2D4D5D6” and “IPAL_D1D2D4D5”. The performance of these runs is given in Table 3.

Table 3: Results of manual visual runs

Rank	run ID	MAP	R-prec
1/10	IPAL_CMP_D1D2D4D5D6	15.96%	19.39%
2/10	IPAL_CMP_D1D2D3D4D5	15.84%	19.22%
3/10	IPAL_CMP_D1D2D3D4D5D6	15.79%	19.62%
4/10	IPAL_D1D2D4D5D6	15.51%	20.58%
5/10	IPAL_cfD1D2D4D5D6	15.50%	20.47%
6/10	IPALcf_D1D2D3D4D5D6	15.20%	20.19%
7/10	IPAL_CMP_D1D2D4D5	14.63%	19.94%
8/10	IPAL_cfD1D2D4D5	14.61%	19.98%
9/10	IPAL_D1D2D4D5	14.61%	19.98%
10/10	IPALcf_D1D2D3D4D5	14.17%	19.57%

From these results, we can find that:

- Applying a probability estimation of each image about its modality and imaging anatomy helps (compare “IPAL_CMP_D1D2D4D5D6” (MAP=0.1596) and “IPAL_cfD1D2D4D5D6” (MAP=0.155));
- Color filtering generally improves performance, but not significantly;
- Using D6 in the combination, the performance is improved. Compare “IPAL_cfD1D2D4D5D6” (MAP=0.155) versus “IPAL_cfD1D2D4D5” (MAP = 0.1461) and “IPAL_D1D2D4D5D6” (MAP=0.1551) versus “IPAL_D1D2D4D5” (MAP=0.1461).

4 UMLS-based Mixed Retrieval

We propose three types of fusion between text and images:

- a simple late fusion (run “IPAL_Cpt_Im”);
- a fusion that uses a visual filtering according to modality concept(s) (runs “IPAL_ModFDT_Cpt_Im”, “IPAL_ModFST_Cpt_Im”, “IPAL_ModFDT_TDF_Im”, and “IPAL_ModFDT_Cpt”);
- an early fusion of UMLS-based visual and textual indexes (run “IPAL_MediSmart.1” and “IPAL_MediSmart.2”).

Table 4: Results of automatic mixed runs

Rank	run ID	MAP	R-prec
1/37	IPAL_Cpt_Im	30.95%	34.59%
2/37	IPAL_ModFDT_Cpt_Im	28.78%	33.52%
3/37	IPAL_ModFST_Cpt_Im	28.45%	33.17%
4/37	IPAL_ModFDT_TDF_Im	27.30%	37.74%
5/37	IPAL_ModFDT_Cpt	27.22%	37.57%
17/37	IPAL_MediSmart_1	6.49%	10.12%
30/37	IPAL_MediSmart_2	4.20%	6.57%

The first fusion method is a late fusion of visual and textual similarity measures. The similarity between a mixed query $Q = (Q_I, Q_T)$ (Q_I : image(s), Q_T : text) and a couple composed of an image and the associated medical report (I, R) is then given by:

$$\lambda(Q, I, R) = \alpha \frac{\lambda_V(Q_I, I)}{\max_{z \in \mathcal{D}_I} \lambda_V(Q_I, z)} + (1 - \alpha) \frac{\lambda_T(Q_T, R)}{\max_{z \in \mathcal{D}_T} \lambda_T(Q_T, z)} \quad (6)$$

where $\lambda_V(Q_I, I)$ denotes the visual similarity between the visual query Q_I and an image I , $\lambda_T(Q_T, R)$ denotes the textual similarity between the textual query Q_T and the medical report R , \mathcal{D}_I denotes the image database, and \mathcal{D}_T denotes the text database. The factor α allows the control of the weight of the textual similarity with respect to the image similarity. After some experimentations on imageCLEFmed 2005, we choose $\alpha = 0.7$. In order to compare similarities in the same range, each similarity is divided by the corresponding maximal similarity value on the entire database. The result of the corresponding run, “IPAL_Cpt_Im”, given in Table 4 show the good complementarity of the visual and textual indexing: from 26% for the textual retrieval and 6% for the visual retrieval, the mixed retrieval provides 31% of MAP. The best results on imageCLEFmed 2006 in terms of MAP and R-precision (i.e. precision after R retrieved images, where R is the number of relevant images) were obtained with this simple late fusion.

The second type of fusion exploits directly the UMLS index of images. Indeed, it is based on a direct matching between concepts extracted from the textual query and conceptual image indexes. This direct matching is done automatically with the use of the Unified Medical Language System. More specifically, a comparison between the query concepts related to modality and the image modality index is done in order to remove all aberrant images. The decision rule is the following: an image I is admissible for a query modality MOD_Q only if:

$$P(\text{MOD}_Q|I) > \tau(\text{MOD}_Q) \quad (7)$$

where $\tau(\text{MOD}_Q)$ is a threshold defined for the modality MOD_Q . This decision rule defines a set of admissible images for a given modality MOD_Q : $\{I \in \mathcal{D}_I : P(\text{MOD}_Q|I) > \tau(\text{MOD}_Q)\}$. The final result is then the intersection of this set and the ordered set of images retrieved by any system. This modality filter is particularly interesting for filtering textual retrieval results. Indeed, several images of different modalities can be associated to the same medical report. The ambiguity is thus removed when using a visual modality filtering. We test this approach with, first, a fixed threshold for all modality $\tau(\text{MOD}_Q) = 0.15$ (“ModFST”) based on experimental tests on 2005 CLEF medical task, and, second, an adaptive threshold for each modality according a confidence degree given to the classifier according this modality (“ModFDT”). The adaptive thresholding performs slightly better than the constant thresholding (compare “IPAL_ModFDT_Cpt_Im” and “IPAL_ModFST_Cpt_Im” in Table 4). In fact, we have over-estimated these thresholds for most modalities. Indeed, when this modality filtering is applied to the late fusion results the results

⁶“IPAL_CMP_D1D2D3D4D5D6” corresponds to “IPAL_CMP_D1D2D3D4D5D” in ImageCLEFmed where the last letter was missing

decrease of 2 points (compare “IPAL_ModFDT_Cpt_Im” and “IPAL_Cpt_Im”). That means that this filtering not only removes aberrant images but also relevant images. This filtering nevertheless increases the results of the purely textual retrieval approach from 26% to 27% (see “IPAL_ModFDT_Cpt” in Table 4 and “IPAL_Textual_CDW” in Table 1). Moreover, this filtering is relevant if the user - which is often the case - is more interested in the precision in the first retrieved images that in the mean average precision. Indeed, the Figure 1 shows that the adaptive modality filtering on the late fusion results (“IPAL_ModFDT_Cpt_Im”) and even directly on the textual results (“IPAL_ModFDT_Cpt”) provides a better precision than the late fusion results (“IPAL_Cpt_Im”) for the first retrieved documents (until 30 when applied on textual results, until 50 when applied on the mixed results).

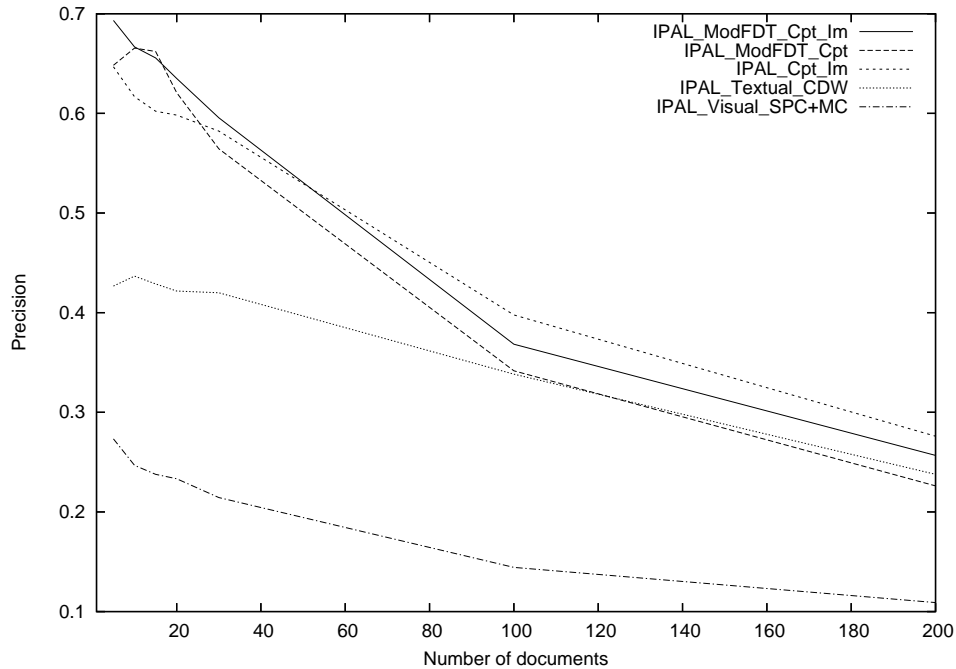


Figure 1: Precision for N retrieved documents on ImageCLEFmed 2006 queries.

We also submitted two runs concerning the early fusion of UMLS-based visual and textual indexes. A Semantic level fuzzyfication algorithm takes into account the frequency, the localization, the confidence and the source of the information [19]. Unfortunately, errors were found after the submission. Using the corrected algorithm, we obtain 24% of mean average precision on ImageCLEFmed 2006. We have to note that the dimension filtering and re-weighting used in the other IPAL mixed runs are not applied here, which explains in part the difference of precision. In fact, this result is higher than the results obtained with runs that do not use this dimension filtering (20% for mixed retrieval, 23% for textual retrieval). We currently develop clustering techniques to improve the retrieval results. A fuzzy min-max boosted K-means clustering approach gives promising results on CASImage database.

5 Conclusion

In this paper, we have proposed a medical image retrieval system that represents both texts and images at a very high semantic level using concepts from the Unified Medical Language System. Textual, visual, and mixed approaches derived from this system were evaluated on ImageCLEFmed 2006. A structured framework was proposed to bridge the semantic gap between low-level image

features and the semantic UMLS concepts. A closed visual system based on manual query construction and visual task fusion was also tested to go as far as possible using a purely visual approach. From the results on ImageCLEF 2006, we can conclude that the textual approaches capture more easily the semantics of the medical queries, providing better results than purely visual retrieval approaches. Indeed, the best visual approach in 2006 - that corresponds to a result of our closed system - only provides 16% of MAP against 26% of MAP for the best textual results. Moreover, the results show the potential of conceptual indexing, especially when using a semantic dimension filtering: we obtained the best textual and mixed results in imageCLEF 2006 using our UMLS-based system. The benefit of working in a fusion framework has been demonstrated. Firstly, visual retrieval results are enhanced by the fusion of global and local similarities. Secondly, mixing textual and visual information improves significantly the system performance. Besides precision in the first documents increases when using a visual modality filtering, allowing 68% of mean precision on the 10 first documents and 62% of mean precision for the 30 first documents on the 30 queries of ImageCLEF 2006. We are currently investigating the potential of an early fusion scheme using appropriate clustering methods. In the near future, we plan to use the LVM terms from local indexing for semantics-based retrieval (i.e. cross-modal retrieval: processing textual query on LVM-based image indexes). A visual filtering based on local information could also be derived from the semantic local indexing.

References

- [1] A. Aronson. Effective mapping of biomedical text to the UMLS metathesaurus: The MetaMap program. In *Proceedings of the Annual Symposium of the American Society for Medical Informatics*, pages 17–21, 2001.
- [2] K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D.M. Blei, and M.I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.
- [3] C.M. Bishop. *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995.
- [4] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximisation and its applications to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1026–1038, 2002.
- [5] Jean-Pierre Chevallet. X-IOTA: An open XML framework for IR experimentation application on multiple weighting scheme tests in a bilingual corpus. *Lecture Notes in Computer Science (LNCS), AIRS'04 Conference, Beijing*, 3211:263–280, 2004.
- [6] W.W. Chu, F. C. Alfonso, and K.T. Ricky. Knowledge-based image retrieval with spatial and temporal constructs. *IEEE Transactions on Knowledge and Data Engineering*, 10:872–888, 1998.
- [7] Paul Clough, Henning Muller, Thomas Desealers, Michael Grubinger, Thomas Lehmann, Jeffery Jensen, and William Hersh. The CLEF 2005 automatic medical image annotation task. Springer Lecture Notes in Computer Science. To appear.
- [8] J.G. Dy, C.E. Brodley, A.C. Kak, L.S. Broderick, and A.M. Aisen. Unsupervised feature selection applied to content-based retrieval of lung images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):373–378, 2003.
- [9] T. Huang, Y. Rui, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proceedings of the IEEE International Conference on Image Processing*, pages 815–818, 1997.
- [10] T. Joachims. *Learning to Classify Text using Support Vector Machines*. Kluwer, 2002.

- [11] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, and Z. Protopapas. Fast and effective retrieval of medical tumor shapes. *IEEE Transactions on Knowledge and Data Engineering*, 10:889–904, 1998.
- [12] T.M. Lehmann et al. Content-based image retrieval in medical applications. *Methods Inf Med*, 43:354–361, 2004.
- [13] J.H. Lim and J.-P. Chevallet. VisMed: a visual vocabulary approach for medical image indexing and retrieval. In *Proceedings of the Asia Information Retrieval Symposium*, pages 84–96, 2005.
- [14] J.H. Lim and J.S. Jin. Discovering recurrent image semantics from class discrimination. *EURASIP Journal of Applied Signal Processing*, 21:1–11, 2006.
- [15] Y. Liu et al. Semantic based biomedical image indexing and retrieval. In L. Shapiro, H.P. Kriegel, and R. Veltkamp, editors, *Trends and Advances in Content-Based Image and Video Retrieval*. Springer, 2004.
- [16] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler. A review of content-based image retrieval systems in medical applications - clinical benefits and future directions. *International Journal of Medical Informatics*, 73:1–23, 2004.
- [17] W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. QBICproject: querying images by content, using color, texture, and shape. In W. Niblack, editor, *Storage and Retrieval for Image and Video Databases*, volume 1908, pages 173–187. SPIE, 1993.
- [18] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. *International Journal of Computer Vision*, 18:233–254, 1996.
- [19] D. Racoceanu, C. Lacoste, R. Teodorescu, and N. Vuilleminot. A semantic fusion approach between medical images and reports using UMLS. In *Proceedings of the Asia Information Retrieval Symposium (Special Session)*, Singapore, 2006.
- [20] Chi-Ren Shyu, Christina Pavlopoulou, Avinash C. Kak, Carla E. Brodley, and Lynn S. Broderick. Using human perceptual categories for content-based retrieval from a medical image database. *Computer Vision and Image Understanding*, 88(3):119–151, 2002.
- [21] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1349–1380, 2000.
- [22] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [23] Xiong Wei, Qiu Bo, Tian Qi, Xu Changsheng, Ong Sim-Heng, and Foong Kelvin. Combining multilevel visual features for medical image retrieval in ImageCLEF 2005. In *Cross Language Evaluation Forum 2005 workshop*, page 73, Vienna, Austria, September 2005.
- [24] T. Westerveld. Image retrieval : Content versus context. In *Recherche d'Information Assistee par Ordinateur*, 2000.
- [25] J. K. Wu, A. Desai Narasimhalu, B.M. Mehtre, C.P. Lam, and Y.J. Gao. CORE: a content-based retrieval engine for multimedia information systems. *Multimedia Systems*, 3:25–41, 1995.
- [26] R. Zhao and W. Grosky. Narrowing the semantic gap - improved text-based web document retrieval using visual features. *IEEE Transactions on Multimedia*, 4(2):189–200, 2002.