

A Semantic Fusion Approach Between Medical Images and Reports Using UMLS

Daniel Racoceanu^{1,2}, Caroline Lacoste¹,
Roxana Teodorescu^{1,3}, and Nicolas Vuilleminot^{1,4}

¹ IPAL-Image Perception, Access and Language - UMI-CNRS 2955
Institute for Infocomm Research, A*STAR, Singapore
{visdaniel, viscl, sturot}@i2r.a-star.edu.sg
<http://www.i2r.a-star.edu.sg/>

² University of Franche-Comte, Besancon, France

³ "Politehnica" University from Timisoara, Romania

⁴ Ecole Nationale Supérieure de Mécaniques et Microtechniques de Besancon, France

Abstract. One of the main challenges in content-based image retrieval still remains to bridge the gap between low-level features and semantic information. In this paper, we present our first results concerning a medical image retrieval approach using a semantic medical image and report indexing within a fusion framework, based on the Unified Medical Language System (*UMLS*) metathesaurus. We propose a structured learning framework based on Support Vector Machines to facilitate modular design and extract medical semantics from images. 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. Visual indexes and textual indexes - extracted from medical reports using *MetaMap* software application - constitute the input of the late fusion module. A weighted vectorial norm fusion algorithm allows the retrieval system to increase its meaningfulness, efficiency and robustness. First results on the CLEF medical database are presented. The important perspectives of this approach in terms of semantic query expansion and data-mining are discussed.

1 Introduction

Many programs and tools have been developed to formulate and execute queries based on the visual content and to help browsing large multimedia repositories. Still, no general breakthrough has been achieved with respect to large varied databases with exogenous documents. Many questions with respect to speed, semantic descriptors or objective image interpretations are still unanswered. In the medical field, digital images are produced in ever-increasing quantities and used for diagnostics and therapy. With digital imaging and communications in medicine (DICOM), a standard for image communication has been set and patient information can be stored with the actual image(s), although still a few problems prevail with respect to the standardization. In several articles, content-based access to medical images for supporting clinical decision-making has been

proposed that would ease the management of clinical data and scenarios for the integration of content-based access methods into picture archiving and communication systems (PACS) have been created [16].

There are several reasons why there is a need for additional, alternative image retrieval methods apart from the steadily growing rate of image production. For the clinical decision making process it can be beneficial or even important to find other images of the same modality, the same anatomic region of the same disease. Although part of this information is normally contained in the DICOM headers and many imaging devices are DICOM-compliant at this time, there are still some problems. DICOM headers contain a high rate of errors: error rates of 16% have been reported by [9] for the field anatomical region. This can hinder the correct retrieval of all wanted images. Clinical decision support techniques such as case-based reasoning [12] or evidence-based medicine [4] can even produce a stronger need to retrieve images that can be valuable for supporting certain diagnoses. It could even be imagined to have Image-Based Reasoning (IBR) as a new discipline for diagnostic aid. Decision support systems in radiology [10] and computer-aided diagnostics for radiological practice as demonstrated at the RSNA (Radiological Society of North America) are on the rise and create a need for powerful data and meta-data management and retrieval [1].

It needs to be stated that the purely visual image queries as they are executed in the computer vision domain will most likely not be able to ever replace text-based methods as there will always be queries for all images of a certain patient, but they have the potential to be a very good complement to text-based search based on their characteristics. Still, the problems and advantages of the technology have to be stressed to obtain acceptance and use of visual and text-based access methods up to their full potential.

Besides diagnostics, teaching and research especially are expected to improve through the use of visual access methods as visually interesting images can be chosen and can actually be found in the existing large repositories. The inclusion of visual features into medical studies is another interesting point for several medical research domains. Visual features do not only allow the retrieval of cases with patients having similar diagnoses but also cases with visual similarity but different diagnoses. In teaching it can help lecturers as well as students to browse educational image repositories and visually inspect the results found.

According to those remarks, this study introduces a semantic indexing approach of the medical report and the medical image, according to the concepts associated to an unified medical modeling system. This approach give a complementary description of the textual and visual characteristic of a medical case, using a balanced weighted vectorial norm fusion method at the conceptual level, by taking into account the confidence, the localization and the frequency of the associated concepts.

The remainder of this paper is organized as follows. Section 2 provides a brief state of the art in the content-based image retrieval, focusing on text and image fusion. In section 3, we introduce the semantic indexing approach, by presenting the main ideas used to close the semantic gap and to develop a complementary

text-image fusion approach. Finally, some operational approaches and results are synthesized in the section 4 in CLEF ⁵ benchmark context, by extracting the main points for the conclusions and the future developments.

2 Brief Overview of the Content-Based Image Retrieval and Fusion Methods

Although access methods in image databases already existed in the beginning of the 1980s [5], the content-based image retrieval (CBIR) started in the 1990s [17],[8] and has been an extremely active research area over the last 10 years [16]. There is growing interest in CBIR because of the limitations inherent in metadata-based systems. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database. Moreover, it is possible to miss images that use different synonyms in their descriptions.

Content-based image retrieval (CBIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. "Content-based" means that the search makes use of the contents of the images themselves, rather than relying on human-imputed metadata such as captions or keywords.

The visual features, used for indexing and retrieval, are classified in [11] into three classes:

- primitive features that are low-level features such as color, shape, and texture;
- logical features that are medium-level features describing the image by a collection of objects and their spatial relationships;
- abstract feature that are semantic/contextual features.

Current CBIR systems generally make use of primitive features [17], [18]. However, some general semantic layers are insufficient to model medical knowledge. Consequently results are poor when common algorithms are general system offers interpretation of images or even medium level concepts as they can easily be captured with text. This loss of information from image data to a representation by features is called the semantic gap [21]. To reduce this semantic gap, specialized retrieval systems have been proposed in literature [11],[20],[6]. Indeed, the more a retrieval application is specialized for a limited domain, the smaller the gap can be made by using domain knowledge.

Some interesting initiative like Image Retrieval in Medical Applications (IRMA) [13] and medGIFT [16] propose a general content-based medical image retrieval system. Even if the results are considerably improved with those systems in the general content-based retrieval framework, still remain the challenge to bridge the semantic gap and the fusion between heterogeneous medical multimedia sources.

⁵ Cross Evaluation Forum Language: <http://www.clef-campaign.org/>

Some recent studies [3], [2] associate explicitly the image and the text. Statistic methods are used for modeling the occurrence of document keywords and visual characteristics. This system is very sensitive to the quality of the segmentation of the images.

Some published works [19] propose to find a semantic of the text using Latent Semantic Analysis. Visual information are extracted using color histograms and edges, and next clustered using the Principal Component Analysis method. In this approach, the two modalities are not combined using the LSA.

Other more recent initiatives study the use of LSA techniques. [22] apply the LSA method to color and texture 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. [23] use the LSA method on the textual and visual combined information. The tests are not really consistent, like the database contain only few documents.

Interesting recent initiatives using LSA for combining text and image are presented in [7] for general images retrieval by the combination of different sources (text and image) at the features level, after clustering. This approach, even promising, is still subject to some empiric thresholds and parameters like the number of visual clusters associated to the image indexes. The database - more consistent than in [23] -, contains nevertheless only 4500 documents.

In our approach, we initiate a semantic level indexing and fusion, according to a well known medical metathesaurus: the Unified Medical Language System. Even if this ontology is still perfectible (some incoherences and inconsistencies still remain), this approach allows us to fuse the medical report and image at the homogeneous medical conceptual level and the tests give us some reliable indicators about its efficiency (the tests are operated on Clef Medical Image Database containing 50 000 medical images). Moreover, the conceptual level indexing and fusion will facilitate all future works concerning the semantic query and case expansion, the context-aware navigation and query and the data-mining.

3 General framework: a unified medical indexing

The main idea of our approach is to take deeply into account existing structured medical ontology/knowledge to reduce the indexing semantic gap and improve the efficiency of the current general CBIR systems for medical images. The semantic indexing framework proposed consists of three main modules:

1. Medical image analysis and conceptualization
2. Medical report analysis and conceptualization
3. Medical conceptual processing and balanced fusion

This article proposes a semantic approach by focusing on the medical image and report fusion. This approach uses concepts from the Unified Medical Language System (UMLS) to index both images and medical reports (Fig. 1). After an introduction to *UMLS* in section 3.1, each module of our system is described in detail in the following sections.

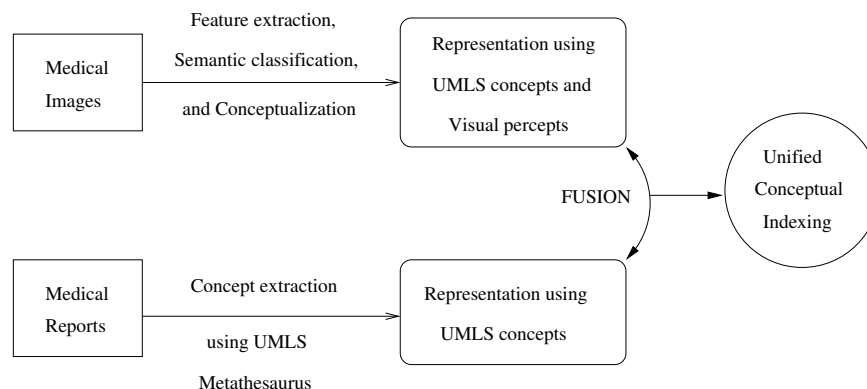


Fig. 1. Unified conceptual indexing

3.1 Use of the UMLS Ontology for improving the Fusion and Retrieval Efficiency

The purpose of NLM's⁶ Unified Medical Language System (UMLS)⁷ is to facilitate the development of computer systems that behave as if they "understand" the meaning of the language of biomedicine and health.

The UMLS Metathesaurus is a very large, multi-purpose, and multi-lingual vocabulary database that contains information about biomedical and health related concepts, their various names, and the relationships among them. The Metathesaurus is organized by concept or meaning. All concepts in the Metathesaurus are assigned to at least one semantic type from the Semantic Network. This provides consistent categorization of all concepts in the Metathesaurus at the relatively general level represented in the Semantic Network.

In order to filter the UMLS concepts and relationships needed for the fusion, medical case semantic indexing and query expansion, the Metathesaurus UMLS Knowledge Source has been used. This Metathesaurus is composed by medical concepts and is distributed with several tools (programs) that facilitate their use, including the MetamorphoSys install and customization program. MetamorphoSys is the UMLS installation wizard and customization tool included in each UMLS release. It may be used to exclude vocabularies that are not required or licensed for use in local applications and to select from a variety of data output options and filters.

Using MetamorphoSys, we have extracted all the Concept Unique Identifiers (CUI) from the UMLS Metathesaurus UMLS file in order to build the concept layer used for medical cases (image+report) indexing.

⁶ US National Library of Medicine - <http://www.nlm.nih.gov/>

⁷ Unified medical Language System - <http://www.nlm.nih.gov/research/umls/>

3.2 Medical image analysis and conceptualization

Concerning the image indexing part, our aim is to associate to each image, or to each image region, a semantic label that corresponds to a combination of UMLS concepts and visual percepts. We define three types of UMLS concepts that could be associated to one image or one region:

- modality concepts that belong to the following UMLS semantic type: “Diagnostic Procedure”;
- anatomy concepts that belong to the following UMLS semantic types: “Body Part, Organ, or Organ Component”, “Body Location or Region”, “Body Space or Junction”, or “Tissue”;
- pathology concepts that belong to the following UMLS semantic types: “Acquired Abnormality”, “Disease or Syndrome”, or “Injury or Poisoning”

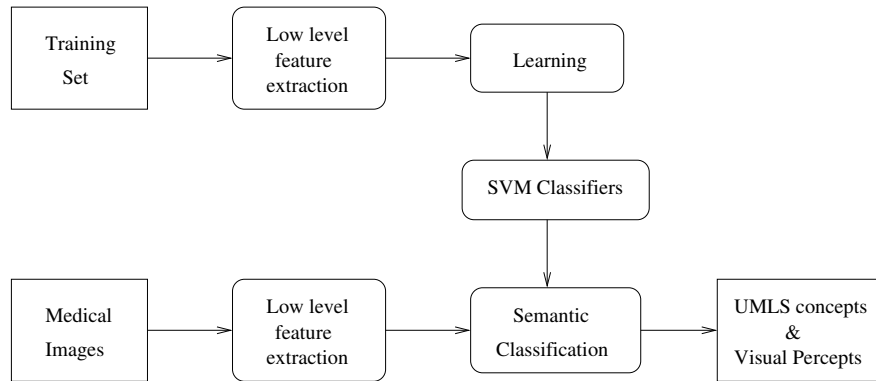


Fig. 2. Conceptual image indexing using SVMs

We propose a structured learning framework based on Support Vector Machines (SVMs) to facilitate modular design and extract medical semantics from images (Fig. 2). 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 one modality concept, one anatomy concept, and, sometimes, to one pathology concepts.

After presenting our general learning framework, we detail both approaches hereafter.

General statistical learning framework: Firstly, a set of disjoint semantic tokens with visual appearance in medical images is selected to define a Visual and Medical vocabulary. This notion of using a visual and semantic vocabulary to represent and index image has been applied to consumer images in [15, 14]. Here, we use UMLS concepts to represent each token in the medical domain.

Secondly, low-level features are extracted from image region instances to represent each token in terms of color, texture, shape, etc.

Thirdly, these low-level features are used as training examples to build a semantic classifier according the Visual and Medical Vocabulary. We use a hierarchical classification based on SVMs. First, a tree whose leaves are the Visual-Medical terms is constructed. The upper levels of the tree consist to auxiliary classes that cluster similar terms with respect to their visual appearance. A learning process is performed at each node in the following way. If a node corresponding to a cluster \mathcal{C} has $N_{\mathcal{C}}$ direct children, $N_{\mathcal{C}}$ SVM classifiers are learned to classify in one class against the $N_{\mathcal{C}} - 1$ other classes. The positive and negative examples for a class $c \in \mathcal{C}$ is respectively given by the instances of the term(s) associated to the class c and the instances of the terms associated to the $N_{\mathcal{C}} - 1$ other classes. This is a conditional learning in a sense as the classifiers are learned given that a class belongs to a given cluster.

The classifier according the Visual and Medical vocabulary is finally designed from the tree of SVM conditional classifiers in the following way. The conditional probability that an example z belong to a class c given that the class belong to the cluster \mathcal{C} is first computed using the softmax function:

$$P(c|z, \mathcal{C}) = \frac{\exp^{\mathcal{D}_c(z)}}{\sum_{j \in \mathcal{C}} \exp^{\mathcal{D}_j(z)}} \quad (1)$$

where \mathcal{D}_c is the signed distance to the SVM hyperplane that separate class c from the other classes of the cluster \mathcal{C} . The probability of a Visual-Medical term vMT_i (i. e. a leaf of the tree) for an example z is finally given by:

$$P(\text{vMT}_i|z) = P(\text{vMT}_i|z, \mathcal{C}^1(\text{vMT}_i)) \prod_{l=1}^{L-1} P(\mathcal{C}^l(\text{vMT}_i)|z, \mathcal{C}^{l+1}(\text{vMT}_i)) \quad (2)$$

where L is the number of hierarchical levels, $\mathcal{C}^1(\text{vMT}_i)$ denotes the cluster to which vMT_i belongs, and $\mathcal{C}^l(\text{vMT}_i)$ the cluster to which $\mathcal{C}^{l-1}(\text{vMT}_i)$ belongs. $\mathcal{C}^L(\text{vMT}_i)$ is the cluster containing all the defined classes (it corresponds to the tree root).

Global UMLS indexing according Modality: The global UMLS indexing is based on a two level hierarchical classifier according modality concepts. This

modality classifier has been learned from 4000 images separated in 32 classes: 22 grey level modalities, and 10 color modalities. This images come from the CLEF database (2500 examples), from the IRMA database (300 examples), and from the web (1200 examples). The training images from CLEF database was obtained from modality concept extraction using medical report. A manual filtering on this extraction to remove irrelevant examples had to be performed. We plan to automatize this filtering in the near future.

The first level corresponds to a classification according 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 HSV three first moments computed on the entire image. The second level corresponds to the classification according 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 filtering 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 is applied to each feature. For each SVM classifier, we adopted a RBF kernel with modified city-block distance between feature vectors y and x that equally takes into account each type of feature:

$$|y - x| = \frac{1}{F} \sum_{f=1}^N \frac{|y^f - x^f|}{N_f} \quad (3)$$

where F is the number of type of features ($F = 1$ for the grey versus color classifier, $F = 3$ for the conditional modality classifiers: color, texture, thumbnails), x_f is the feature vector of type f , and $x = \{x_1, \dots, x_F\}$.

The probability of a modality MOD_i for an image z is given by equation 2. More precisely, we have:

$$P(\text{MOD}_i|z) = \begin{cases} P(\text{MOD}_i|z, C)P(C|z) & \text{if } \text{MOD}_i \in C \\ P(\text{MOD}_i|z, G)P(G|z) & \text{if } \text{MOD}_i \in G \end{cases} \quad (4)$$

where C and G respectively denote the color and the grey level clusters.

A modality concept can thus be assigned to an image z using the following formula:

$$L(z) = \max_i P(\text{MOD}_i|z) \quad (5)$$

The classifier has been first learned on the half on the training dataset to evaluate its performance in the other half of the training dataset. The error rate on the test set is about 18%, with recall and precision rates larger than 70% for a large majority of the classes. The classification is quite good given the intra-variability of some classes with respect to the class inter-variability. For example, differentiate a brain MRI and a brain CT can be a hard task, even for a human operator.

After learning (using the entire learning set), each database image is indexed according to modality given its low-level features. The index values are the probability values given by equation (4).

Local UMLS indexing: To better capture the medical image content, we propose to extend this first modeling for local patch classification in local visual and semantic tokens (LocVisMed terms). Each LocVisMed term is expressed as a combination of Unified Medical Language System (UMLS) concepts from Modality, Anatomy, and Pathology UMLS semantic types. In these experiments, we have adopted color and texture features from patches (i. e. small image blocks) and a non-hierarchical classifier based on SVMs and the softmax function given by equation (1). A Semantic Patch Classifier was finally designed to classify a patch according to 64 LocVisMed terms. 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 filtering using 5 scales and 6 orientations. Zero-mean normalization is applied to both the color and texture features. We adopted a RBF kernel with modified city-block distance given by equation (3).

The training set is composed of 3631 patches extracted from images mostly coming from the web (1182 images coming from the web and 158 images from the CLEF collection). The classifier has been first learned on a first half of this training set to be evaluated on a second half. The error rate of this classifier is about 30%.

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

3.3 Medical report analysis and conceptualization

In order to improve conventional text Information Retrieval approaches, we pass from the text syntactic level to the semantic one. In the medical field, this step requires the use of specialized concepts from available thesaurus and metathesaurus. In this sense, the Unified Medical Language System helps us to acquire this pertinent higher level indexing, using a specific *UMLS* concepts extractor like *MetaMap*, provided by the National Library of Medicine (NLM). A concept is then an abstraction of this synonymous set of terms. Thus, conceptual text indexing consists of associating a set of concepts to a document and uses

it as index. This set of concepts should cover the document theme. Conceptual indexing naturally solves the term mismatch and the multilingual problem.

3.4 Medical conceptual processing and balanced fusion

Even if from the medical point of view, a medical case can have more than one associated medical image, for the retrieval purpose, we consider one case c composed by one medical report and one medical image.

The main idea of our approach is to use the medical rapport and medical image conceptual and/or visual-concept indexing in order to build an homogeneous high level fusion approach for improve the retrieval system performances.

Such a medical case c will bring thus an indexing from the associated image and respectively medical report:

$$A_{txt}^c = \begin{bmatrix} CUI_{i_1} & \lambda_{txt_{i_1}}^c \\ \dots & \dots \\ CUI_{i_n} & \lambda_{txt_{i_n}}^c \end{bmatrix}, \quad A_{img}^c = \begin{bmatrix} CUI_{j_1} & \lambda_{img_{j_1}}^c \\ \dots & \dots \\ CUI_{j_m} & \lambda_{img_{j_m}}^c \\ VC_{k_1} & \lambda_{img_{vc_{k_1}}}^c \\ \dots & \dots \\ VC_{k_p} & \lambda_{img_{vc_{k_p}}}^c \end{bmatrix} \quad (6)$$

with: $\lambda_{txt_{i_l}}^c = \mu_{txt_{i_l}}^c \cdot \nu_{txt_{i_l}}^c \cdot \omega_{txt_{i_l}}^c \cdot \gamma_{txt_{i_l}}^c, l = 1, \dots, n$
 $\lambda_{img_{j_q}}^c = \mu_{img_{j_q}}^c \cdot \nu_{img_{j_q}}^c \cdot \omega_{img_{j_q}}^c \cdot \gamma_{img_{j_q}}^c, q = 1, \dots, m$
 $\lambda_{img_{vc_{k_s}}}^c = \mu_{img_{vc_{k_s}}}^c \cdot \nu_{img_{vc_{k_s}}}^c \cdot \omega_{img_{vc_{k_s}}}^c \cdot \gamma_{img_{vc_{k_s}}}^c, s = 1, \dots, p$

where:

- CUI is the notation for an UMLS concepts,
- VC are visual concepts, mix between UMLS concepts and perception concepts,
- μ means the indexing confidence degree,
- ν is associated to the local relative frequency of the concept,
- ω corresponds to the spatial localization fuzzy weight,
- γ represents the semantic tree (modality, anatomy, biology, pathology and direction) fuzzy weight.

For the medical case c (medical image and associated medical report document) and the corresponding UMLS extracted concept CUI_i , we obtain the next fuzzy confidence indexing vector:

$$\lambda_i^c = [\lambda_{txt_i}^c, \lambda_{img_i}^c] = [\mu_{txt_i}^c \cdot \nu_{txt_i}^c \cdot \omega_{txt_i}^c \cdot \gamma_{txt_i}^c, \mu_{img_i}^c \cdot \nu_{img_i}^c \cdot \omega_{img_i}^c \cdot \gamma_{img_i}^c] \quad (7)$$

If the concept is a visual concept VC_s , we have:

$$\lambda_{vc_{s}}^c = [0, \mu_{img_{vc_{s}}}^c \cdot \nu_{img_{vc_{s}}}^c \cdot \omega_{img_{vc_{s}}}^c \cdot \gamma_{img_{vc_{s}}}^c] \quad (8)$$

As the semantic medical report and image use the same homogeneous ontology, the fusion takes into account the norm of the so obtained global credibility vector λ_i^c as a projection of the c medical case to each concept CUI_i :

$$pr_{CUI_i}(c) = \|\lambda_i^c\|$$

Obviously, we will find the same kind of projection of the case c to each associated visual concept VC_s of the corresponding medical image:

$$pr_{VC_s}(c) = \|\lambda_{vc-s}^c\|$$

A medical case c will be then characterized by the vector Λ_c of the global credibilities of all associated *UMLS* and *visual* concepts:

$$\Lambda_c = \left[\|\lambda_{i_1}^c\|, \|\lambda_{i_2}^c\|, \dots, \|\lambda_{i_n}^c\|, \|\lambda_{vc_1}^c\|, \dots, \|\lambda_{vc_p}^c\| \right]^T$$

The parameters used for the projection of a medical case on *CUI* or *VC* are obtained by direct image and text pre-computation (indexing) - e.g. fuzzy indexing confidence degree μ and the local relative frequency ν - or by fuzzyfication - for the spatial localization ω and semantic tree membership γ fuzzy weights.

The indexing confidence degree μ is a fuzzy result directly given by the classifiers (equation (4) used for the medical image indexing. For the text pre-processing, a text indexing software (in our case, *MetaMap*) has been used for calculate the local relative frequency $\nu_{txt_i}^c$ of the concept occurrence in the given medical report. If a patch extraction method is used for the image, the local relative frequency $\nu_{img_i}^c$ will be computed using the relative weight of this concept versus all the patches of the image. The spatial localization ω corresponds - for the text indexing - to the importance of the medical report XML tag or the section from which the concept comes. For example, the *< Diagnosis >* paragraph of the medical report, the physician synthesized the most important keywords describing the disease (pathology) and the anatomic part. This tag will thus be more important than (par example) the *< Description >* tag. In order to fuzzy this subjective parameter, we propose the fuzzy membership sets presented in the Fig. 3.

For the image indexing, the spatial localization corresponds to a special weight accorded to a particular place in the image (e.g. the central patches for a CT medical image or a circle sector shape object for a Doppler indexing).

The semantic tree membership γ intent to give a particular weight to a concept belonging to a particular semantic tree (modality, anatomy, pathology, biology, direction) according to the indexing source (image or text) of this concept. Par example, a modality concept coming from the medical image indexing will have more importance that a modality extracted from the medical report. Opposite, a pathology concept will have more "confidence" if extracted by the medical report indexing service (Fig. 4), with:

$$\begin{aligned} \text{if } CUI_i \in MOD &\Rightarrow \alpha = \sigma \\ \text{if } CUI_i \in ANA/BIO &\Rightarrow \alpha = 2\sigma \\ \text{if } CUI_i \in PAT &\Rightarrow \alpha = 3\sigma \end{aligned}$$

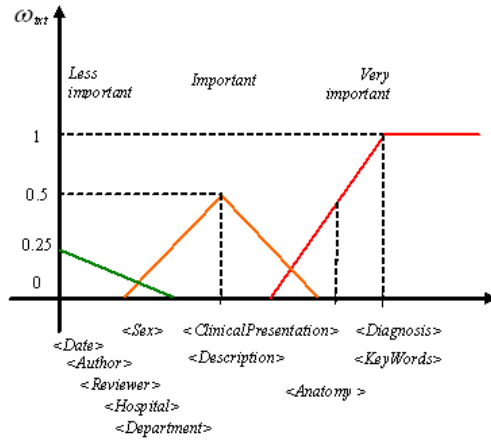


Fig. 3. Spatial localization parameter ω_{txt} fuzzyfication

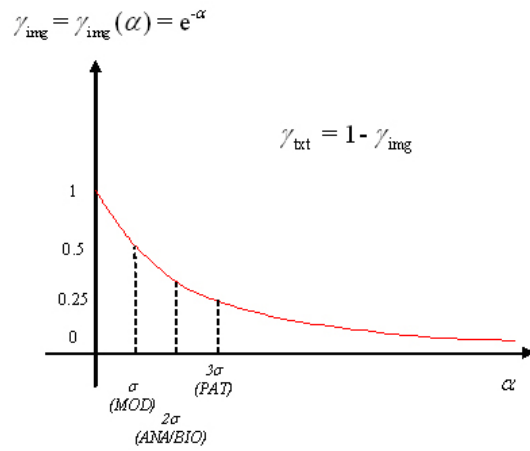


Fig. 4. Semantic tree membership γ fuzzyfication

where σ corresponds to the size of the fuzzy membership function associated to the influence zone of each semantic type.

4 Results on CLEF 2005 Medical Images Benchmark

We apply our approach on the medical image collection of CLEF⁸ Cross Language Image Retrieval track. This database consists of four public datasets (CASImage⁹, MIR¹⁰, PathoPic¹¹, PEIR¹²) containing 50000 medical images with the associated medical report in three different languages. In 2005, 134 runs were evaluated on 25 queries containing at least one of the following axes: anatomy (ex: heart), modality (ex: X-ray), pathology or disease (ex: pneumonia), abnormal visual observation (ex: enlarged heart).

We test five approaches on these 25 queries to evaluate the benefit of using a UMLS indexing, especially in a fusion framework. First, three UMLS image indexing were tested on the visual queries:

1. Global UMLS image indexing presented in section 3.2 / Retrieval based on the Manhattan distance between two modality indexes (modality probabilities);
2. Local UMLS image indexing presented in section 3.2 / Retrieval based on the Manhattan distances between LocVisMed histograms.
3. Late fusion of the two visual indexing approaches (1) and (2)

The two last tests respectively concerns textual indexing and retrieval using UMLS and the fusion between this textual indexing and the Global UMLS image indexing (i.e. modality indexing). The integration of local information is under developmen. The text indexing uses *MetaMap* software and give us the *UMLS* associated concepts *CUI* with the corresponding relative frequencies ν . The medical image indexing uses a global image approach giving essentially *UMLS* modalities concepts *CUI_{MOD}*, the associated frequency ν and SVM confidence degree μ .

With these partial indexing information, we build the global confidence degree λ such as:

$$\lambda_i^c = [\lambda_{txt\ i}^c\ \lambda_{img\ i}^c] = [1 \cdot \nu_{txt_i}^c\ \gamma_{txt_i}^c\ \mu_{img_i}^c\ \nu_{img_i}^c\ \gamma_{img_i}^c]$$

Comparative results are given in table 1. For the visual indexing and retrieval without textual information, our results are quite good with respect to the best 2005 results, especially when local and global UMLS indexes are mixed. Our textual approach is on the first 2005 results (between 9% and 20%) but

⁸ CLEF - Cross Language Evaluation Forum - www.clef-campaign.org/

⁹ <http://www.casimage.com>

¹⁰ <http://gamma.wustl.edu/home.html>

¹¹ <http://alf3.urz.unibas.ch/pathopic/intro.html>

¹² <http://peir.path.uab.edu>

significantly under the approach proposed by Jean-Pierre Chevallet (IPAL) in 2005 that uses a textual filtering on MeSH terms according three dimensions: Modality, Anatomy, and Pathology. The high average precision is principally due to this textual filtering. We have to notice that the association between MeSH terms and a dimension had to be done manually. With the use to UMLS metha-thesaurus, we have the advantage to have access to these dimension thanks to the semantic type associated to each UMLS concept.

We can verify the benefit of the fusion between image and text : from 16% for the text only and 10% for the image only, we reach the 24% with a mixed indexing and retrieval. We are slightly superior in average precision than the mixed approaches that do not use dimension filtering.

Table 1. Comparative results on the medical task of CLEF 2005

Method	Visual	Textual	MAP
Fusion between UMLS image indexes	X		12.11%
Global UMLS image indexing	X		10.38%
Local UMLS image indexing	X		06.56%
Best automatic visual run in 2005 (GIFT)	X		09.42%
UMLS textual indexing		X	16.41%
Best automatic textual run in 2005 (DFMT) Dimension Filtering on MESH Terms		X	20.84%
UMLS mixed indexing	X	X	24.07 %
Best automatic mixed run in 2005 (DFMT) + local semantic indexing	X	X	28.21 %
Best automatic mixed run in 2005 without dimension filtering	X	X	23.89 %

5 Conclusion

The presented researches constitutes a very promising approach of semantic indexing and late balanced fusion, in the medical cases retrieval framework.

One important contribution of this study is the use of a web-based up to date medical metathesaurus (UMLS) in order to "standardize" the semantic indexing of the text and the medical image. This allows to work on the same level for both medical media and to have thus an homogeneous complementary point of view about the medical case.

The introduction of a late semantic fusion for each common UMLS concept, using multiple criteria weighted norm involving the frequency of a concept, the indexing confidence degree, the spatial localization weight and the semantic tree belonging, constitutes another important point to be underlined. The so obtained

vectorial norm represents a balanced projection of the medical case (document and image) to the given UMLS concept, a consistent information enabling a reliable and robust semantic retrieval.

Future developments become very promising using this homogeneous balanced semantic fusion. Appropriate clustering methods should be able to bridge to medical multimedia data-mining, opening the way to the evidence-based medicine and other advanced medical research applications and studies.

In future evolutions of this application, our fusion approach will be improved using the local visual information derived from the proposed local patch classifier. Indeed, this method is complementary to the global medical image analysis and will certainly improve the global retrieval results.

Otherwise, the use of the incremental learning based on the initial database clustering, should be able to facilitate the development of an efficient real-time medical case-based reasoning.

Finally, a semantic query and case expansion policy can be deployed using the symbolic and statistic relation available in *UMLS* at the first time, and the contextual behavior information extracted from the real use of the retrieval system in the second time.

References

1. H. Abe, H. and MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, and K. Doi. Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996-2001 rsna scientific assemblies. *RadioGraphics*, 23(1):255–265, 2003.
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. K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 408–415, 2001.
4. A. A. T. Bui, R. K. Taira, J. D. N. Dionision, D. R. Aberle, S. El-Saden, and H. Kangarloo. Evidence-based radiology. *Academic Radiology*, 9(6):662–669, 2002.
5. N.-S. Chang and K.-S. Fu. Query-by-pictorial-example. *IEEE Transactions on Software Engineering*, 6(6):519–524, 1980.
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. C. Fleury. Apports reciproques des informations textuelles et visuelles par analyse de la semantique latente pour la recherche d'information. Master's thesis, Intelligence, Interaction and Information, CLIPS-MRIM Grenoble, France, IPAL UMI CNRS 2955, Singapore, June 2006.
8. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: The qbic system. *IEEE Computer*, 28(9):23–32, 1995.
9. M. O. Guld, M. Kohnen, D. Keyzers, H. Schubert, B. B. Wein, J. Bredno, and T. M. Lehmann. Quality of dicom header information for image categorization.

- In *Proceedings of the International Symposium on Medical Imaging*, volume 4685, pages 280–287, San Diego, CA, USA, 2002. SPIE.
10. C. E. Kahn. Artificial intelligence in radiology: Decision support systems. *Radio-Graphics*, 14:849–861, 1994.
 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. C. LeBozec, M.-C. Jaulent, E. Zapletal, and P. Degoulet. Unified modeling language and design of a case-based retrieval system in medical imaging. In *Proceedings of the Annual Symposium of the American Society for Medical Informatics*, Nashville, TN, USA, 1998.
 13. T. M. Lehmann, M.O. Gld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, and B. B. Wein. Content-based image retrieval in medical application. *Methods of Information in Medicine*, 43(4):354–361, 2004.
 14. J. 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.
 15. J. Lim and J. Jin. A structured learning framework for content-based image indexing and visual query. *Multimedia Systems*, 10:317–331, 2005.
 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. S. Sclaroff, M. la Cascia, and S. Sethi. Unifyng textual and visual cues for content-based image retrieval on the world wide web. *Computer Vision and Image Understanding*, 75(1/2):86–98, 1998.
 20. 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. *Computer Vision and Image Understanding*, 75:111–132, 1999. Special issue on content-based access for image and video libraries.
 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. T. Westerveld. Image retrieval : Content versus context. In *Recherche d'Information Assistee par Ordinateur*, 2000.
 23. 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.