

Semantic Retrieval of Radiological Images with Relevance Feedback

Camille Kurtz¹(✉), Paul-André Idoux¹, Avinash Thangali², Florence Cloppet¹,
Christopher F. Beaulieu², and Daniel L. Rubin²

¹ LIPADE, University Paris Descartes, Paris, France
`camille.kurtz@parisdescartes.fr`

² Department of Radiology, School of Medicine, Stanford University, Stanford, USA

Abstract. Content-based image retrieval can assist radiologists by finding similar images in databases as a means to providing decision support. In general, images are indexed using low-level features, and given a new query image, a distance function is used to find the best matches in the feature space. However, using low-level features to capture the appearance of diseases in images is challenging and the semantic gap between these features and the high-level visual concepts in radiology may impair the system performance. In addition, the results of these systems are fixed and cannot be updated based on user's intention. We present a new framework that enables retrieving similar images based on high-level semantic image annotations and user feedback. In this framework, database images are automatically annotated with semantic terms. Image retrieval is then performed by computing the similarity between image annotations using a new similarity measure, which takes into account both image-based and ontological inter-term similarities. Finally, a relevance feedback mechanism allows the user to iteratively mark the returned answers, informing which images are relevant according to the query. This information is used to infer user-defined inter-term similarities that are then injected in the image similarity measure to produce a new set of retrieved images. We validated this approach for the retrieval of liver lesions from CT images and annotated with terms of the RadLex ontology.

Keywords: Image retrieval · Riesz wavelets · Image annotation · RadLex · Semantic gap · Relevance feedback · Computed tomographic (CT) images

1 Introduction

Diagnostic radiologists are now confronted with the challenge of efficiently interpreting cross-sectional studies that often contain thousands of images [1]. A promising approach to maintain interpretative accuracy in this “deluge” of data is to integrate computer-based assistance into the image interpretation process. Content-based image retrieval (CBIR) approaches could assist users in

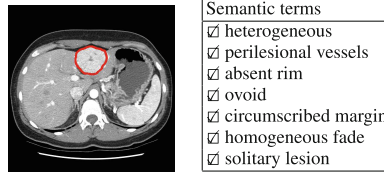


Fig. 1. A lesion of the liver (in red) in CT image annotated with semantic terms (Color figure online).

finding visually similar images within large image collections. This is usually performed by example, where a query image is given as input and an appropriate distance is used to find the best matches in the corresponding feature space [2]. CBIR approaches could then provide real-time decision support to radiologists by showing them similar images with associated diagnoses.

Under CBIR models, images are generally indexed using imaging features extracted from regions of interest (ROI) of the images (*e.g.*, lesions) and focus on their contents (*e.g.*, shape, texture). Although these low-level features are powerful to automatically describe images, they are often not specific enough to capture subtle radiological concepts in images (semantic gap). Despite many efforts conducted to integrate more robust features (*e.g.*, “bag-of-visual-words” [3–5]) into CBIR systems, their performances are often limited by the features low-level properties because they cannot efficiently model the user’s visual observations and semantic understanding [6]. Since this problem remains unsolved, current research in CBIR focuses on new methods to characterize the image with higher levels of semantics, closer to that familiar to the user [7].

In recent work on medical image retrieval with semantics, the images were characterized using terms from ontologies [8]. These terms, which are linked to the user’s high-level understanding of images (Fig. 1), can be used to describe accurately image content (*e.g.*, lesion shape, enhancement). Since terms describe the image contents using the terminology used by radiologists during their observations, they can be considered as powerful features for CBIR [9]. In general, images are represented as vectors of values where each element represents the likelihood of appearance of a term, and the similarity between images is evaluated by computing the vector distance. However, two issues remain unsolved when using terms to characterize medical images. A first issue is the automation of image annotation: usually the terms are manually provided by radiologists. Although many approaches have been proposed to predict these semantic features from computational ones [10], this automation remains challenging for complex lesions. A second issue is that most of the existing systems based on semantic features do not consider the intrinsic relations among the terms for retrieving similar images, and they treat each semantic feature as totally independent of the others.

To deal with this double issue, we proposed in [11] a semantic framework that enables retrieval of similar images based on their visual and semantic properties.

In this framework, database images are annotated and indexed with semantic terms contained in ontologies, which are automatically predicted from the image content using robust Riesz texture features. Given this high-level image representation, the retrieval of similar images is performed by computing the similarity between image annotations using a new measure, which takes into account both image-based and ontological inter-term similarities. The combination of these two strategies provides a means of accurately retrieving similar images in databases based on image annotations and can be considered as a potential solution to the semantic gap problem. As a limit, the results provided by these systems are fixed and cannot be directly updated by the user based on his ultimate intention. A strategy employed in CBIR systems to obtain a better approximation of the user's expectations and preferences is the relevance feedback (RF). It allows the user to weight the returned answers, informing then the retrieval system which images are relevant according to a given image query [12]. Thus, RF is a real-time learning strategy that adapts the answer from a retrieval system exploring the user interaction.

Usually the core scenario of considering the CBIR process as a RF technique can be summarized as follows. For a given query image, the CBIR system retrieves in the image database an initial set of results, ranked according to a pre-defined similarity metric. The user provides judgment on the current retrieval, as to whether the proposed samples are correct or wrong, and possibly to what degree [13]. The system learns from the feedback, and provides a new set of results which are then submitted to the user approval. The system loops until the user is satisfied with the result set whether convergence is possible. As the system captures the user's intention when a new query is performed, the resulting set of images can be continually improved until the gain flattens, according to the iterative learning process [14]. Regarding the learning step of the RF technique, several algorithms with different approaches have been proposed so far [15, 16]. These approaches can be divided into two main categories: The "query point movement" techniques consider that a query is represented by a single query center in the feature space. Therefore, at each user interaction cycle, the strategy estimates an ideal query center in the query space, moving the query center towards the relevant examples and away from the irrelevant ones. On the other hand, the "re-weighting techniques" usually focus on adjusting weights to each dimension of the feature vector emphasizing some dimensions and diminishing the influence of others.

Based on these considerations, we propose an extension of our semantic image retrieval framework based on the user feedback. Instead of considering a classical RF strategy to update the query center in the feature space or to adjust the weights to each dimension of the query feature vector, we use RF to capture similarities and correlations between the terms employed to describe the image contents. Specifically, given a new query image, our motivation is to use the feedback from the user, regarding which retrieved images are relevant or not, to learn inter-term similarities based on user judgment. This information is used to infer a set of "user-defined" inter-term similarities that are then injected in the

image similarity measure to produce a new set of retrieved images. This process is repeated until the retrieval results contain enough relevant images to satisfy the user needs. This real-time learning strategy modifies the similarity measure used to retrieve similar images based on semantic annotations and adapts the answer from a retrieval system in accordance to the radiologist's expectations.

This article is organized as follows. Section 2 first recalls our framework for the retrieval of images annotated with semantic terms (Sect. 2.1). The remainder of this section then presents the proposed extension based on the user feedback (Sect. 2.2). This novel framework is then evaluated in the context of the retrieval of liver lesions extracted from CT images (Sect. 3). Conclusions and perspectives are then presented (Sec. 4).

2 Methodology

2.1 Semantic Retrieval of Radiological Images [11]

The original workflow of the our semantic CBIR framework [11] is divided into four steps that can be grouped in two phases (Fig. 2-① to ④):

- An *offline* phase (2 steps) is used to build a visual model of the terms employed to characterize the images. Step 1 consists of learning, from the database

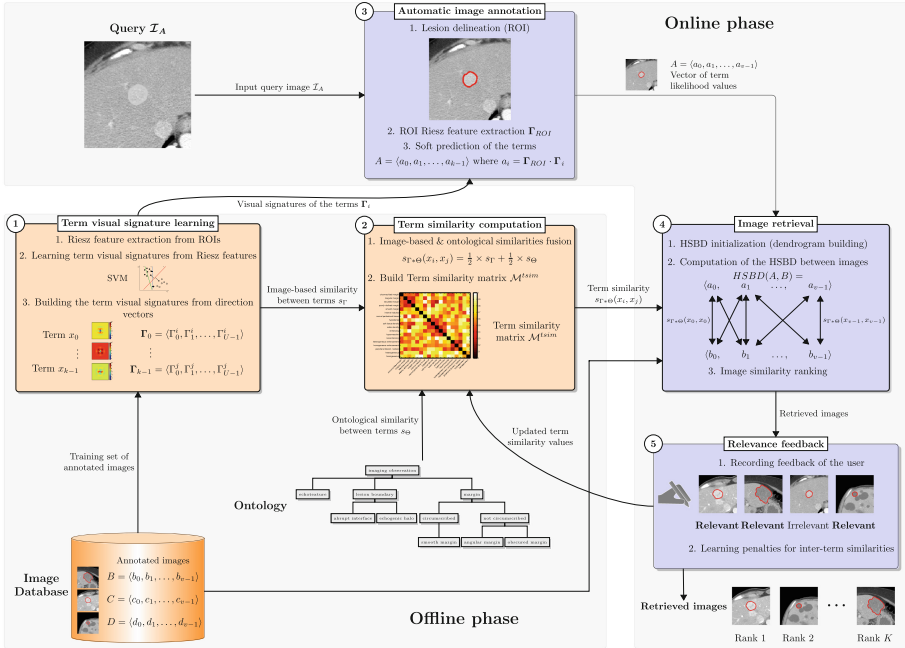


Fig. 2. Workflow of the proposed semantic framework for medical image retrieval.

images, a visual signature for each ontological term based on Riesz wavelets. These texture signatures are used both to predict the image annotations and to establish visual “image-based” similarities between the terms. Step 2 consists of pre-computing the global term similarity values using a combination of their image-based and ontological relations;

- An *online* phase (2 steps) is used to retrieve similar images in the database given a query image. Step 3 consists of automatically annotating this image by predicting term likelihood values based on the term signatures built offline. Step 4 consists of comparing the query to previously annotated images by computing the distance between their term likelihood vectors. Vectors are compared using the hierarchical semantic-based distance (HSBD) [17], which enables to consider the term similarities computed offline.

Offline Phase

Step 1. Learning of the Term Visual Signatures. We use an automatic strategy to predict terms belonging to an ontology Θ that characterize the lesion contents. This strategy, originally proposed in [18], relies on the automatical learning of the term visual signatures from texture features derived from the image ROIs (Fig. 2-①).

To reduce the semantic space search, we created pre-defined lists of terms taken from an ontology Θ . These terms are used to describe the image contents in a specific application. Among these terms, we selected those describing the margin and internal texture of the lesions, since these are key aspects that describe the appearance of lesions. We denote as $\mathcal{X} = \{x_0, x_1, \dots, x_{k-1}\}$ with $x_i \in \Theta$ this vocabulary.

Given a training set of previously annotated image ROIs, this approach learns the image description of each term using support vector machines (SVM) and Riesz wavelets. Each annotated ROI is divided in a set of 12×12 image patches extracted from the lesion margin and internal texture. Each patch is characterized by the energies of multi-scale Riesz wavelets and a gray-level intensity histogram and then represents an instance in the feature space. The learning step relies on SVMs, which are used to build term visual signatures in this feature space. The direction vector of the maximal separating hyperplane in one-versus-all configurations defines the term signature. Once the signatures have been learned, we obtain for each term a model that characterizes its visual description in the image. The visual signature of a term $x_i \in \mathcal{X}$ can be modeled as the direction vector $\Gamma_i = \langle \Gamma_0^i, \Gamma_1^i, \dots, \Gamma_{U-1}^i \rangle$ where each Γ_u^i models the weight of the u -th Riesz template. The term models are used both to predict the presence likelihood of the terms for new image ROIs and to establish the texture similarities between terms.

Step 2. Term Similarity Assessment. The image retrieval step takes into account the term relations when comparing images described by vectors of terms. We proposed to compute the term similarities using both their image-based and ontological relations.

To model the similarity between the k terms of the considered vocabulary $\mathcal{X}$, we define a $k \times k$ symmetric term similarity matrix $\mathcal{M}^{tsim}$ that contains the

intrinsic relations between all the k terms of $\mathcal{X}$. To fill this matrix, we use a similarity function $s_{\Gamma*\Theta}$ based on the combination of the image-based similarity s_Γ provided by the visual signatures of the terms and the semantic similarity s_Θ extracted from the ontological structure Θ .

Image-Based Term Similarity: Let x_i and x_j be two semantic terms. The image-based similarity between two terms x_i, x_j can be evaluated by computing the Euclidean distance between their visual signatures $\mathbf{F}_i, \mathbf{F}_j$ as $s_\Gamma(x_i, x_j) = \sqrt{\sum_{u=0}^{U-1} |\Gamma_u^i - \Gamma_u^j|^2} \cdot \frac{1}{\omega_{norm}^\Gamma}$ where ω_{norm}^Γ is a normalization factor. This similarity models the proximity between the terms according to their image textural appearance.

Semantic Term Similarity: The semantic similarity between two terms x_i, x_j belonging to an ontology can be evaluated by considering edge-based measures that consist of inferring the semantic similarity between terms from the ontology structure [19].

We define $path(x_i, x_j) = \{l_0, \dots, l_{n-1}\}$ as a set of links connecting x_i and x_j in the ontology. In order to quantify a semantic similarity between x_i and x_j , an intuitive method has been proposed in [20]. It relies on a cluster-based strategy that combines the minimum path length between the semantic terms and the taxonomical depth of the considered branches. The underlying idea is that the longer is the path, the more semantically distant the terms are. Starting from the root of the hierarchy, this measure requires the creation of clusters for each main branch of the ontology (each branch is considered as a cluster of term nodes). The idea is then to assess the common specificity (CS) of two terms by subtracting the depth of their lowest common ancestor (LCA) from the depth D_c of their common cluster. The CS is used to consider that lower level pairs of term nodes are more similar than higher level pairs. We recently extended this definition [21] to normalize it and to give an equal weight to the path length and the common specificity features: $s_\Theta(x_i, x_j) = \log \left(\sqrt{\min_{\forall p} [path_p(x_i, x_j)] \cdot CS(x_i, x_j)} + \gamma \right) \cdot \frac{1}{\omega_{norm}^\Theta}$ where ω_{norm}^Θ is a normalization factor evaluating the maximal similarity value between two terms and $CS(x_i, x_j) = D_c - depth(LCA(x_i, x_j))$ is the CS of the terms.

Combination of Image-Based and Semantic Similarities: To combine the image-based and the semantic similarities, we define a weighted sum as $s_{\Gamma*\Theta}(x_i, x_j) = \frac{1}{2} \cdot s_\Gamma(x_i, x_j) + \frac{1}{2} \cdot s_\Theta(x_i, x_j)$ that considers equally the texture and ontological similarities between terms.

Online Phase

Step 3. Automatic Annotation of a Query Image. Let $\mathcal{I}_A$ be a query image. A lesion in the query image $\mathcal{I}_A$ is first delineated to capture the boundary of a ROI. The next step is to automatically characterize the ROI visual content in terms of respective likelihoods of semantic terms belonging to $\mathcal{X}$.

The visual signatures $\mathbf{F}_i$ learned offline for each term $x_i \in \mathcal{X}$ are used to automatically annotate the content of the ROI of the query image $\mathcal{I}_A$. The ROI instance is expressed in terms of the energies E_u of the multi-scale u -th Riesz

templates as $\mathbf{I}_{ROI} = \langle E_0, E_1, \dots, E_{U-1} \rangle$. The likelihood value $a_i \in [0, 1]$ of each term x_i is computed as the dot product between the ROI instance $\mathbf{I}_{ROI}$ and the respective visual signatures $\mathbf{I}_i$. Once the query image $\mathcal{I}_A$ has been “softly” annotated, a vector of semantic features can be built as $A = \langle a_0, a_1, \dots, a_{k-1} \rangle$. It constitutes a synthetic representation of $\mathcal{I}_A$, which forms the feature clue for retrieval purpose.

Step 4. Image Retrieval with Term Similarities. Once the query image $\mathcal{I}_A$ has been characterized with a vector of semantic features, this image description can be used to retrieve similar images in the database based on their vector distances. To this end, the hierarchical semantic-based distance (HSBD) [17, 22] was extended to enable the comparison of vectors of semantic features based on term similarities.

The computation of HSBD relies on the iterative merging of the semantically closest vector elements (*i.e.*, terms) to create coarser vectors of higher semantic levels. The order of fusion between the terms is determined by building a dendrogram¹ (Fig. 4) from the term similarity matrix $\mathcal{M}^{sim}$ (built offline) that contains the combination of image-based and semantic similarities $s_{\Gamma * \Theta}$ between all the k terms of $\mathcal{X}$. After each iteration, the Manhattan distance is computed between the couple of (coarser) vectors created previously. The resulting series of distances enables assignment of vector similarities at different semantic levels. The distances belonging to this series are then fused to provide the HSBD _{$s_{\Gamma * \Theta}$} distance value.

2.2 Extension of the Semantic CBIR Framework Based on User Feedback

As a limit of this system, the retrieval results provided by this CBIR framework are “fixed” and cannot be directly updated by the user based on his ultimate intention. To deal with this issue, we present hereinafter an extension of this system based on relevance feedback. Instead of considering a classical RF strategy to update the query center in the feature space or to adjust the weights to each dimension of the query feature vector, we use RF to capture similarities and correlations between the semantic terms employed to describe the images. Given a query image annotated with a set of terms, our hypothesis is that if an image retrieved by the system has been marked as relevant by the user, it means that this image is described with terms that can be (potentially) considered as semantically “similar” with the terms describing the query image. On the contrary, if an image retrieved by the system has been marked as irrelevant, it means that this image is described with terms that can be considered as semantically “dissimilar” with the terms employed to describe the query image. Based on this assumption, we added to the previous semantic CBIR framework a RF step (Step 5, see Fig. 2-⑤) to learn inter-term similarities based on user

¹ A dendrogram is built using the Ascendant Hierarchical Clustering (AHC) algorithm [23].

judgment. These “user-defined” inter-term similarities are then injected in the image similarity measure to produce a new set of retrieved images.

Let $\mathcal{I}_A$ be a query image automatically annotated with a vector of semantic features $A = \langle a_0, a_1, \dots, a_{k-1} \rangle$. As described earlier, the HSBD distance is used to retrieve similar images in the database (see Step 4) by taking into account a term similarity matrix $\mathcal{M}^{tsim}$ (built offline) that contains the initial combination of image-based and semantic similarities $s_{\Gamma*\Theta}$ between all the k terms of $\mathcal{X}$. The retrieved images are then presented to the user by dissimilarity ranking and the user provides his feedback by selecting a set of relevant images and a set of irrelevant images. Given these image sets, our goal is to learn from them user-defined inter-term similarities in order to update the initial term similarities according to the user judgment. In this preliminary study, we considered a simple penalty algorithm as learning procedure. For each term x_i of the query image belonging to a particular category (see Table 1) and having a high probability of appearance a_i ($a_i > \tau$), we look in the set of relevant results, the images annotated with terms belonging to the same category. For each one of these terms, we select the terms x_j having a probability of appearance a_j comparable to the one of a_i . Once these terms have been identified, we decrease the term dissimilarity value $s_{\Gamma*\Theta}(x_i, x_j)$ (stored in $\mathcal{M}^{tsim}$) with a penalty ρ . The same algorithm is considered with the set of irrelevant images by increasing the initial term dissimilarity value with a penalty ρ . In this preliminary study, we empirically fixed the penalty parameter to $\rho = 0.1$.

As a result, we obtain a new term similarity matrix $\mathcal{M}_{RF}^{tsim}$ modeling the term similarity according to the (1) initial image-based and ontological similarities between terms and the (2) “user-defined” inter-term similarities. Based on this new matrix of term similarities (and the induced dendrogram, see Step 4 and Fig. 4), the HSBD distance is then recomputed ($\text{HSBD}_{s_{\Gamma*\Theta}^{RF}}$) between the query image and the indexed databases images to provide a new set of retrieval results to the user which are then submitted to the user approval. The system loops until the user is satisfied with the result set whether convergence is possible.

3 Experimental Study: Retrieval of Liver Lesions from CT Scans

3.1 Experiments

To assess our framework, we applied it in a system for retrieving liver lesions from a database of 2D CT images. Liver lesions stem from a variety of diseases, each with different visual manifestations. Our database was composed of 72 CT images of liver in the portal venous phase (Fig. 3), including 6 types of lesion diagnoses (Cyst (# 21 images), Metastasis (# 24 images), Hemangioma (# 13 images), Hepatocellular carcinoma (# 6 images), Focal nodular hyperplasia (# 5 images) and Abscess (# 3 images)). We have implemented the proposed semantic framework in a JAVA package extending LIRE (Lucene Image Retrieval) which is an open source library for content based image retrieval [24]. Based

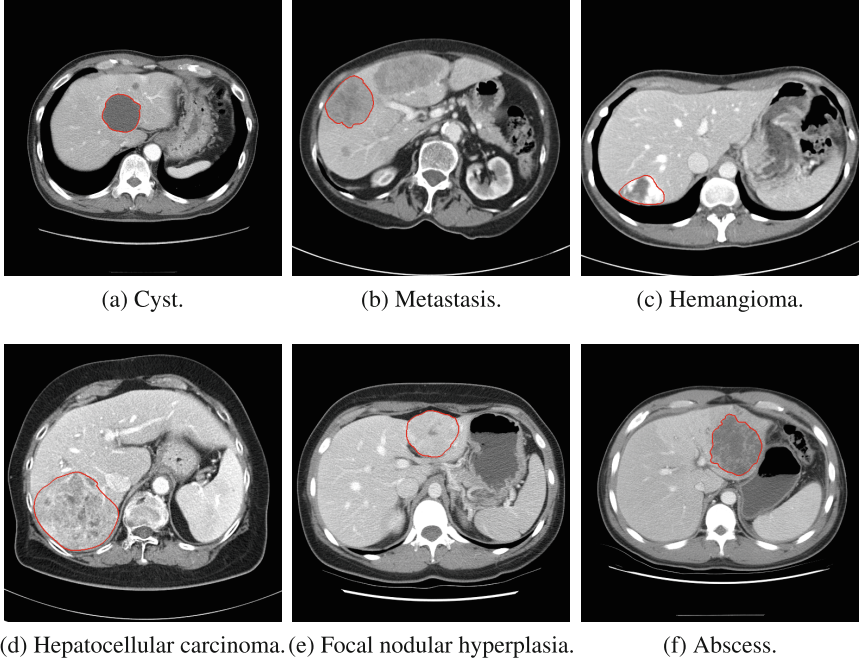


Fig. 3. 6 CT images of liver lesions (boundaries in red) in the portal venous phase (Color figure online).

on the initial retrieval results, we re-ran the system, this time with user input regarding feedback, which initiated the system’s relevance feedback capability. Such a system can be used by radiologists to query the database to find similar medical cases based on the image contents and their opinion.

Our approach requires that lesions on CT images be delineated by a 2D ROI. In this study, a radiologist drew a ROI around the lesion on these images leading to 72 individual ROIs that were used as input to our semantic CBIR framework. Starting from a training set of manually annotated images, the visual signature models of the terms were learned offline using a leave one patient out cross-validation strategy and then used to automatically annotate the 72 ROIs. To build the training set, each lesion was annotated by a radiologist with a set of 18 potential semantic terms (Table 1) from the RadLex ontology [25]. These terms are commonly used by radiologists to describe the lesion margin and the internal texture. In parallel, the offline phase was used to compute the initial term similarity values stored in a 18×18 similarity matrix $\mathcal{M}^{tsim}$.

During the online phase, we selected randomly 10 images as query and ranked the remaining ones according to the initial $HSBD_{s_{\Gamma \times \Theta}}$ distance. In order to evaluate the effect of relevance feedback to refine the query results, we then performed three successive rounds of relevance feedback for each query image. We quantitatively assessed the retrieval performance at each round by comparing

Table 1. RadLex terms used to describe the margin and the internal textures of liver lesions from CT scans. The ontology tree associated with these semantic terms is available online at: <http://bioportal.bioontology.org/ontologies/RADLEX>.

Category	Semantic term
lesion margin and contour	circumscribed margin
	irregular margin
	lobulated margin
	poorly-defined margin
	smooth margin
lesion substance	internal nodules
perilesional tissue	normal perilesional tissue
lesion attenuation	hypodense
	soft tissue density
	water density
overall lesion enhancement	enhancing
	hypervascular
	nonenhancing
spatial pattern of enhancement	heterogeneous enh.
	homogeneous enh.
	peripheral nodular enh.
lesion uniformity	heterogeneous
	homogeneous
lesion effect on liver	abuts capsule of liver

the ranking results obtained with our system to a ranking of reference, which was built from a similarity reference standard (defined for 25×25 image pairs) by two experienced radiologists [8]. We used normalized discounted cumulative gain (NDCG, [26]) to evaluate performance. The NDCG index is used to measure the usefulness (gain) on a scale of 0 to 1 of K retrieved lesions on the basis of their positions in the ranked list compared with their similarity to the query lesion according to a separate reference standard. For each query image, the mean NDCG value was computed at each $K = 1, \dots, 25$. This enables to evaluate the relevance of the results for different number of retrieved images.

3.2 Results

During each round of the relevance feedback phase, the number of images marked by the user as relevant and irrelevant was equal to 6 in average (3 relevant images and 3 irrelevant images).

For visualization purpose, Fig. 4 presents the evolution of the dendrogram structure modeling the merging order of the semantic terms (built according

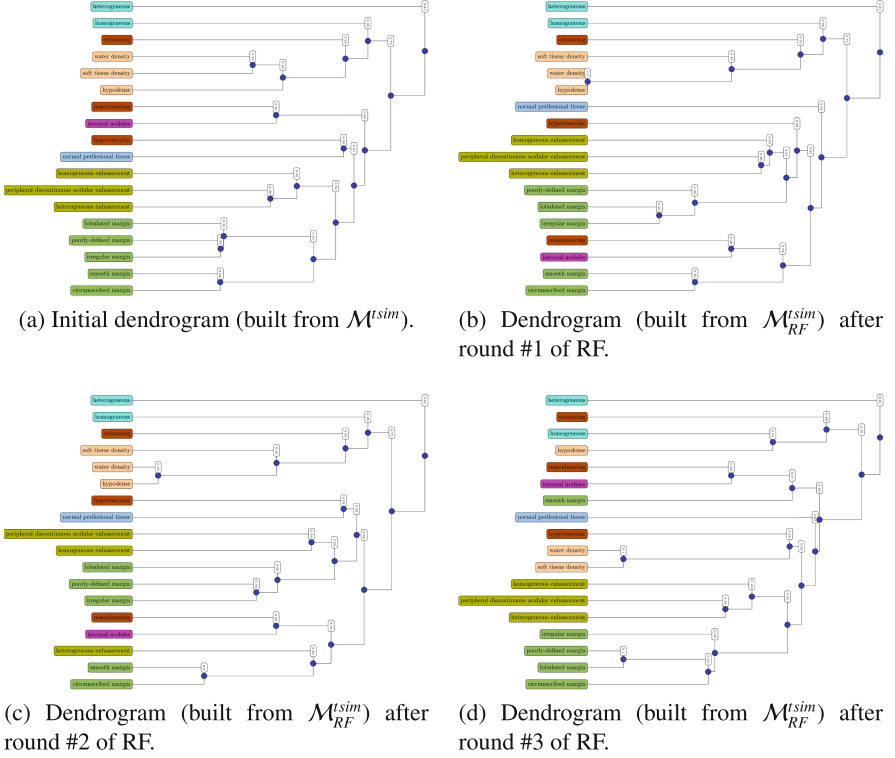


Fig. 4. Evolution of the dendrogram modeling the merging order of the semantic terms (presented in Table 1) after 3 successive rounds of relevance feedback.

to the term similarity matrix $\mathcal{M}_{RF}^{tsim}$, see Sect. 2.2) after 3 successive rounds of relevance feedback from a particular query image. From this figure, one can note that the proposed RF approach enables to capture the user feedback to progressively update and refine the original semantic term similarities (computed from image-based and ontological term similarities) based on the user opinion. For example, the similarities between the semantic terms belonging to the category “lesion margin and contour” (depicted in dark green in the dendrograms) are progressively updated, starting from an initial strong similarity between the terms “poorly-defined margin” and “irregular margin” (Fig. 4 (a)) to a final strong similarity between the terms “poorly-defined margin” and “lobulated margin” (Fig. 4 (d)), reflecting then the user intention.

Figure 5 shows the NDCG scores obtained for 6 (illustrative examples) among the 10 considered image queries and the 3 successive relevance feedback rounds. From these graphs, one can note that the initial results provided by the system with $\text{HSBD}_{s_{\Gamma * \Theta}}$ without the relevance feedback led to NDCG scores already high with a NDCG score higher than 0.62 for all values of K . For 5 experiments, the successive rounds of relevance feedback enable to progressively increase the

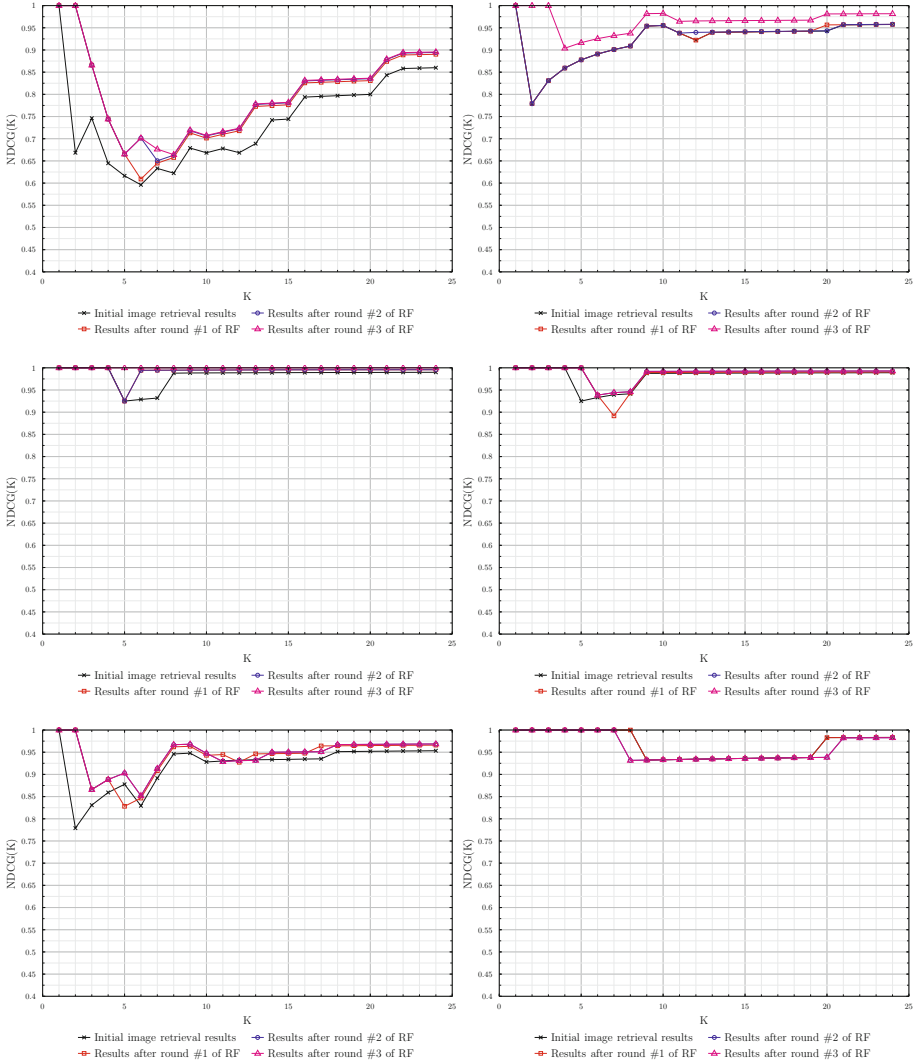


Fig. 5. Image retrieval results for the dataset of CT images of the liver: NDCG scores (before and after rounds of RF) for 6 different experiments.

number of retrieved images in agreement with the user ranking of reference and consequently, increasing the NDCG score values (see for example the graph on top right of Fig. 5 with NDCG score higher than 0.91 for all values of K after the third round of RF). For one experiment (bottom right of Fig. 5) the successive rounds of relevance feedback led to NDCG scores slightly lower than the ones obtained with the initial retrieval results, potentially due to a problem of over-learning of the term semantic similarities.

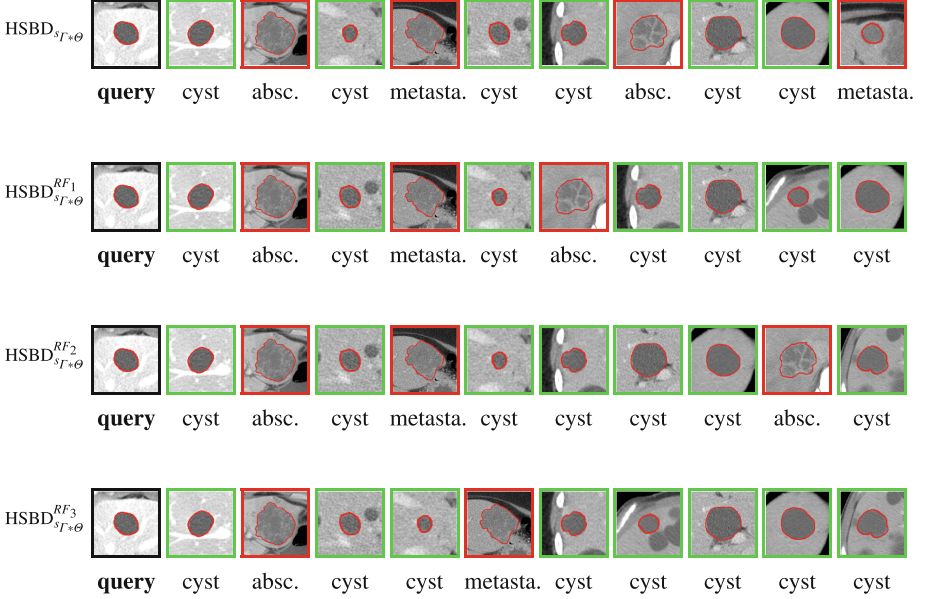


Fig. 6. Examples of image retrieval for a cyst query after the successive rounds of relevance feedback. Dissimilarity rankings go from lowest (left) to highest (right). The query image is surrounded by a black frame. Retrieved images having the same diagnosis as the query image are surrounded by green frames while retrieved images having a different diagnosis are surrounded by red frames (Color figure online).

For qualitative evaluation purpose, we also evaluated the ability of our system to retrieve images of the same diagnosis in the database of $M = 72$ lesions belonging to six types. Figure 6 illustrates retrieval results after the successive rounds of relevance feedback by using a cyst query. Perfect retrieval (in terms of the diagnoses) would result in a ranked list of images with only cyst lesions. From this figure, we observe that the successive rounds of relevance feedback enable to continuously increase the number of cyst lesions retrieved by the system and to take away in the retrieval results positions lesions with a different diagnosis (see for example the abscess lesion in position 6 in row 2, and then in position 9 in row 3).

4 Conclusions and Perspectives

We presented an extension based on relevance feedback of a semantic framework that enables retrieving similar images characterized with semantic annotations. Given an initial set of retrieved images, our extension consists of considering user feedback to capture similarities and correlations between the terms employed to describe the image contents and then to inject them in the image similarity measure to produce a new set of retrieved images. This process is repeated until

the retrieval results contain enough relevant images to satisfy the user needs. A unique aspect of our approach is the consideration of user judgment to update texture-based and semantic-based similarities between terms that describe the image contents when retrieving similar images.

This preliminary work has some limitations. A first limit is directly linked to the presented relevance feedback strategy: Since the RF rounds gradually modify the initial semantic term similarities (obtained from texture-based and semantic-based relations) used to determine image similarity, there is a potential risk of over-fitting. In addition, we did not evaluate quantitatively in this study the updated / learned semantic term similarities. Another limitation is that the dataset and the reference standard used in the image ranking task were small and contained only 25 CT images. We were unable to develop a larger dataset since it is time consuming for radiologists to assess and annotate images with semantic terms.

We plan to enhance the current framework by considering initial semantic term similarities extracted from multiple biomedical ontologies and complementary quantitative imaging descriptors. Another perspective will rely on studying the convergence of the successive relevance feedback rounds. In case of over-fitting, we envisage to develop a strategy to prune back our final learned semantic term similarities after multiple RF rounds to obtain a better trade-off of learning new semantic relations vs. over-fitting. In the future, the presented image-based and user-based term similarity measures could be used to learn or update existing ontologies. Finally, we also plan to involve this system into larger clinical studies and to extend it to deal with 3D CT volumes.

References

1. Rubin, G.D.: Data explosion: the challenge of multidetector-row CT. *Eur. J. Radiol.* **36**(2), 74–80 (2000)
2. Aigrain, P., Zhang, H., Petkovic, D.: Content-based representation and retrieval of visual media: a state-of-the-art review. *Multimedia Tools Appl.* **3**, 179–202 (1996)
3. Van Gemert, J.C., Veenman, C.J., Smeulders, A.W.M., Geusebroek, J.M.: Visual word ambiguity. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(7), 1271–1283 (2010)
4. Yang, W., Lu, Z., Yu, M., Huang, M., Feng, Q., Chen, W.: Content-based retrieval of focal liver lesions using Bag-of-Visual-Words representations of single- and multiphase contrast-enhanced CT images. *J. Digit. Imaging* **25**, 708–719 (2012)
5. Andre, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N.: Learning semantic and visual similarity for endomicroscopy video retrieval. *IEEE Trans. Med. Imaging* **31**(6), 1276–1288 (2012)
6. Mojsilovic, A., Rogowitz, B.: Capturing image semantics with low-level descriptors. In: *IEEE ICIP*, pp. 18–21 (2001)
7. Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-fei, L.: ImageNet: a large-scale hierarchical image database. In: *CVPR*, pp. 248–255 (2009)
8. Napel, S.A., Beaulieu, C.F., Rodriguez, C., Cui, J., Xu, J., Gupta, A., Korenblum, D., Greenspan, H., Ma, Y., Rubin, D.L.: Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results. *Radiology* **256**(1), 243–252 (2010)

9. Ma, H., Zhu, J., Lyu, M.R.T., King, I.: Bridging the semantic gap between images and tags. *IEEE Trans. Multimedia* **12**(5), 462–473 (2010)
10. Zhang, D., Islam, M.M., Lu, G.: A review on automatic image annotation techniques. *Pattern Recogn.* **45**(1), 346–362 (2012)
11. Kurtz, C., Depeursinge, A., Napel, S., Beaulieu, C.F., Rubin, D.L.: On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. *Med. Image Anal.* **18**(7), 1082–1100 (2014)
12. Zhou, X.S., Huang, T.S.: Relevance feedback in image retrieval: a comprehensive review. *Multimedia Syst.* **8**(6), 536–544 (2003)
13. Cheng, P.C., Chien, B.C., Ke, H.R., Yang, W.P.: A two-level relevance feedback mechanism for image retrieval. *Expert Syst. Appl.* **34**(3), 2193–2200 (2008)
14. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. *Pattern Recogn.* **40**(1), 262–282 (2007)
15. Zhang, H., Chen, Z., Li, M., Su, Z.: Relevance feedback and learning in content-based image search. *World Wide Web* **6**(2), 131–155 (2003)
16. Doulamis, N., Doulamis, A.: Evaluation of relevance feedback schemes in content-based in retrieval systems. *Sig. Process. Image Commun.* **21**(4), 334–357 (2006)
17. Kurtz, C., Gañarski, P., Passat, N., Puissant, A.: A hierarchical semantic-based distance for nominal histogram comparison. *Data Knowl. Eng.* **87**(1), 206–225 (2013)
18. Depeursinge, A., Kurtz, C., Beaulieu, C.F., Rubin, D.L.: Predicting visual semantic descriptive terms from radiological image data: preliminary results with liver lesions in CT. *IEEE Trans. Med. Imaging* **33**(8), 1669–1676 (2014)
19. Wu, Z., Palmer, M.: Verbs semantics and lexical selection. In: *ACL*, pp. 133–138 (1994)
20. Al-Mubaid, H., Nguyen, H.A.: A cluster-based approach for semantic similarity in the biomedical domain. In: *IEEE EMBC*, pp. 2713–2717 (2006)
21. Kurtz, C., Beaulieu, C.F., Napel, S., Rubin, D.L.: A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. *J. Biomed. Inf.* **49**(1), 227–244 (2014)
22. Kurtz, C., Passat, N., Gañarski, P., Puissant, A.: A histogram semantic-based distance for multiresolution image classification. In: *IEEE ICIP*, pp. 1157–1160 (2012)
23. Ward, J.H.: Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**(301), 236–244 (1963)
24. Lux, M.: Content based image retrieval with LIRE. In: *ACM MM*, pp. 735–738 (2011)
25. Langlotz, C.P.: RadLex: a new method for indexing online educational materials. *RadioGraphics* **26**(6), 1595–1597 (2006)
26. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **20**(4), 422–446 (2002)

Multimodal Retrieval in the Medical Domain

First International Workshop, MRMD 2015, Vienna,

Austria, March 29, 2015, Revised Selected Papers

Müller, H.; Jimenez del Toro, O.A.; Hanbury, A.; Langs, G.;

Foncubierta Rodriguez, A. (Eds.)

2015, VIII, 153 p. 35 illus. in color., Softcover

ISBN: 978-3-319-24470-9