A SEMANTIC FRAMEWORK FOR THE RETRIEVAL OF SIMILAR RADIOLOGICAL IMAGES BASED ON MEDICAL ANNOTATIONS

Camille Kurtz
LIPADE (EA 2517), Univ. Paris Descartes (France)
camille.kurtz@parisdescartes.fr

Adrien Depeursinge, Christopher F. Beaulieu and Daniel L. Rubin
Dpt. of Radiology, School of Medicine, Stanford Univ. (USA)
{adepeurs,beaulieu,dlrubin}@stanford.edu

ABSTRACT

Image retrieval approaches can assist radiologists by finding similar images in databases as a means to providing decision support. In general, images are indexed using low-level imaging features, and 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. We present a semantic framework that enables retrieving similar images based on high-level semantic image annotations. This framework relies on (1) an automatic approach to predict the annotations as semantic terms from Riesz texture image features and (2) a distance function to compare images considering both texture-based and radiodensity-based similarities among image annotations. Experiments performed on CT images emphasize the relevance of this framework.

Index Terms— Image retrieval, Riesz wavelets, image annotation, RadLex, 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 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 recent efforts conducted to integrate more robust features (*e.g.*, "bag-of-visual-words" [3]) into CBIR systems, their performances are often limited by the low-level properties of these features because they cannot efficiently model the user's high-level visual observations and semantic understanding [4]. 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 [5].

In recent work on medical image retrieval with semantics, the images were characterized using a set of ontological terms [6]. These terms, which are linked to the user's high-level understanding of images, can be used to describe accurately the 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 systems [7]. 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 distance between these vectors. 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 [8], this automation remains challenging for complex lesions. A second issue is that most of the existing CBIR systems based on semantic features do not consider the intrinsic relations (*e.g.*, visual, semantic) among the terms for retrieving similar images, and they treat each semantic feature as totally independent of the others.

We proposed recently [9] a semantic framework that enables retrieval of similar images based on their visual and semantic properties. It relies on two main strategies: (1) an approach to predict the image annotations as ontological terms from Riesz texture features; (2) a distance function to evaluate similarity of image-pairs that considers both the visual and ontological relations among the terms describing the images.

We propose in Sec. 2 an extension of this framework to compare images by considering both the texture-based and the radiodensity-based similarities between the terms describing the images. This combination provides a means of accurately retrieving similar database images that can be considered as a potential solution to reduce the semantic gap. This novel framework is then evaluated in the context of the retrieval of liver lesions extracted from CT images (Sec. 3). Conclusions and perspectives are then presented (Sec. 4).

2. METHODOLOGY

The workflow of the proposed CBIR framework is divided into four steps that can be grouped in two phases (Fig. 2):

– 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 images, a visual signature for each ontological term based on Riesz wavelets. These signatures are used both to predict the image annotations and to establish visual "image-based" similarities between the terms. Step 2 consists of computing term similarities using a fusion of their texture-based and radiodensity-based similarities.

– 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) [10, 11], which enables to consider the term similarities computed offline.

2.1. Offline phase

Step 1. Learning of the term visual signatures. In this framework, we use an automated strategy to predict terms belonging to an ontology Θ that characterize the lesion contents. This strategy, originally proposed in [12, 13], relies on automated learning of the term visual signatures from textural features derived from the image ROIs (Fig. 2-①).

To reduce the semantic space search, we created predefined lists of terms taken from a 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 $\mathcal{X} = \{x_0, x_1, \ldots, x_{k-1}\}$ 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 histogram of the intensity (*i.e.*, radiodensity) in Hounsfield units (HU) in $[-60, 220]$ (with 20 bins) that models the distribution of the gray-levels in the patch. This patch represents an instance in the feature space. The learning step relies then 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 a visual description of the term in the image. The visual signature of a term $x_i \in \mathcal{X}$ can be modeled as the direction vector[1] $\mathbf{\Gamma}_i = \langle \Gamma_0^i, \Gamma_1^i, \ldots, \Gamma_{U-1}^i \rangle$ where each Γ_u^i models the weight of the u-th Riesz template. The same approach is used to learn the importance of each bin of the histogram. Each visual signature $\mathbf{\Gamma}_i$ goes along a 20 bins histogram $\mathbf{\Psi}_i = \langle \Psi_0^i, \Psi_1^i, \ldots, \Psi_{19}^i \rangle$ where each Ψ_u^i models the weight of the u-th bin in HU. Fig. 1 shows 3 visual signatures learned for terms used to annotate liver lesions in CT scans.

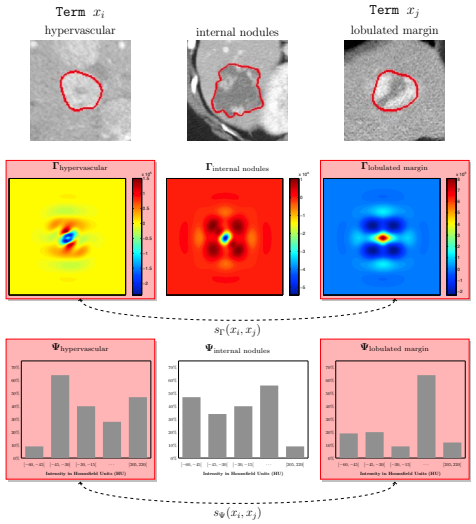


Fig. 1: Visual signatures learned as linear combinations of Riesz wavelets for 3 terms. The color scale shows the signature profiles $\mathbf{\Gamma}_i$ obtained as a weighted sum of Riesz templates. Each signature $\mathbf{\Gamma}_i$ is presented with (1) a ROI where the term modeled by the signature has a high appearance probability and (2) its companion 20 bins histogram $\mathbf{\Psi}_i$.

The term models are used both to predict the presence likelihood of the terms for new image ROIs and to establish the texture and radiodensity similarities between terms.

Step 2. Term similarity assessment. The image retrieval step of this framework takes into account the term relations when comparing images described by vectors of terms. We propose in this work to compute the term similarities using their texture-based and radiodensity-based similarities.

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 $s_{\Gamma * \Psi}$ based on the combination of texture-based s_Γ and radiodensity-based s_Ψ term similarity provided by the visual signatures of the terms.

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

[1]The length $U = J \cdot (N + 1)$ of this vector depends on the order of the Riesz transform N and the number of dyadic scales J (in practice we set $N = 8$ and $J = 3$).

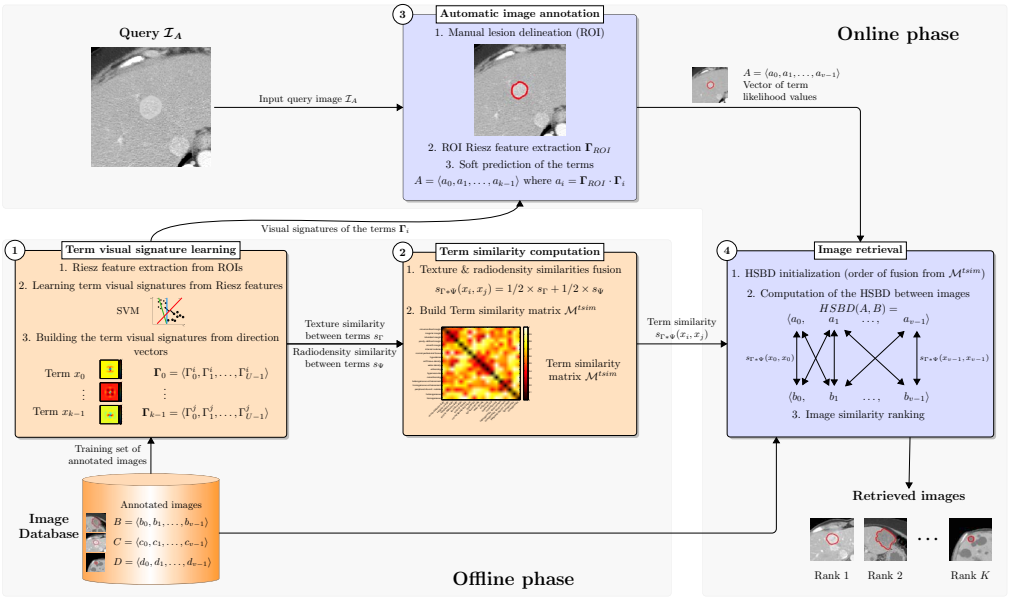


Fig. 2: Workflow of the proposed framework for image retrieval (orange boxes = offline steps; blue boxes = online steps).

Radiodensity-based term similarity: The radiodensity-based similarity between two terms x_i, x_j can be evaluated by computing the Euclidean distance between their histogram signatures $\mathbf{\Psi}_i, \mathbf{\Psi}_j$ as $s_\Psi(x_i, x_j) = \frac{\sqrt{\sum_{u=0}^{19} |\Psi_u^i - \Psi_u^j|^2}}{\omega_{norm}^\Psi}$ where ω_{norm}^Ψ is a normalization factor. This similarity models the proximity between the terms according to their radiodensity appearance in the image (Fig. 1).

Combination of texture and radiodensity similarities: To combine the image-based and the radiodensity-based similarities, we define a weighted sum as $s_{\Gamma*\Psi}(x_i, x_j) = 1/2 \cdot s_\Gamma(x_i, x_j) + 1/2 \cdot s_\Psi(x_i, x_j)$ that considers equally the texture and radiodensity similarities between terms.

2.2. 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 manually delineated to capture the boundary of a ROI. The next step is to characterize the ROI content in terms of respective likelihoods of semantic terms belonging to $\mathcal{X}$.

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

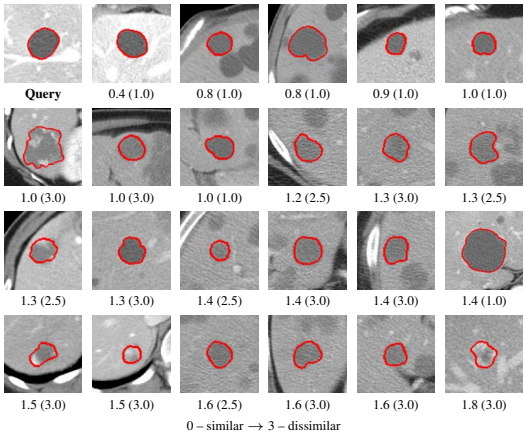
Step 4. Image retrieval with term similarities. Once the query image $\mathcal{I}_A$ has been characterized with a vector of semantic features, this image description can be used to retrieve similar images in the database based on their vector distances. To this end, the hierarchical semantic-based distance (HSBD) [10, 11] 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 from the term similarity matrix $\mathcal{M}^{tsim}$ (built offline) that contains the combination of texture-based and radiodensity-based similarities $s_{\Gamma*\Psi}$ 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 $\text{HSBD}_{s_{\Gamma*\Psi}}$ distance value.

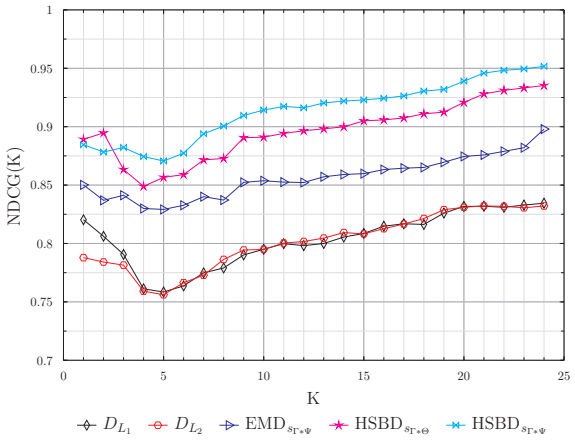
3. LIVER LESIONS RETRIEVAL FROM CT SCANS

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, including 6 types of lesion diagnoses (Cyst, Metastasis, Hemangioma, Hepatocellular carcinoma, Focal nodular hyperplasia and Abscess). We have used the proposed semantic framework in a system for ranking image similarity to a query image. Such a system can be used by radiologists to query the database to find similar medical cases based on the image contents.



(a) Images retrieved for a cyst query. For each image, the distance value computed with $\text{HSBD}_{s_{\Gamma*\Psi}}$ (and the reference similarity value) is presented.



(b) NDCG scores of the five distances evaluated.

Fig. 3: Image retrieval results for the dataset of CT images of the liver.

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 image retrieval 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 (*e.g.*, smooth margin, irregular margin, internal nodules, hypodense, heterogeneous, hypervascular) from the RadLex ontology [14]. 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 term similarity values that were stored in a 18×18 term similarity matrix $\mathcal{M}^{tsim}$.

During the online phase, we withheld each database image and ranked the remaining ones according to $\text{HSBD}_{s_{\Gamma*\Psi}}$ (Fig. 3 (a)). We evaluated the retrieval performance by comparing 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 confirmed radiologists [6]. We used normalized discounted cumulative gain (NDCG, [15]) 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 $K = 1, \ldots, 25$. This enables to evaluate the relevance of the results for different number of retrieved images.

We compared the retrieval results obtained with $\text{HSBD}_{s_{\Gamma*\Psi}}$ to the results obtained by using other existing distances: the Manhattan D_{L_1} and Euclidean D_{L_2} distances, which do no take into account the relations among the terms, the earth mover's distance ($\text{EMD}_{s_{\Gamma*\Psi}}$) [16], which was extended to consider the term relations contained in the $\mathcal{M}^{tsim}$ term similarity matrix built offline, and the previous version of $\text{HSBD}_{s_{\Gamma*\Theta}}$ [9], which evaluates the term relations based on their texture similarity and ontological proximity.

3.2. Results

Fig. 3 (b) shows the NDCG scores obtained for the five considered distances. From this graph, one can note that the D_{L_1} and D_{L_2} distances appeared to yield the worst overall results, with mean case retrieval accuracy equals to 0.80. The $\text{HSBD}_{s_{\Gamma*\Psi}}$ distance appeared to yield the best overall results, with mean case retrieval accuracy equals to 0.92. Results obtained with the $\text{EMD}_{s_{\Gamma*\Psi}}$ and $\text{HSBD}_{s_{\Gamma*\Theta}}$ distances yielded intermediate overall results, with mean case retrieval accuracy equals to 0.85 and 0.89 respectively.

4. CONCLUSIONS AND PERSPECTIVES

We present a semantic framework that enables retrieving similar images based on semantic annotations. These annotations consist of ontological terms, automatically predicted from the image content, which ensure the performance reproducibility with radiologists. A unique aspect of our approach is the consideration of both texture-based and radiodensity-based similarities between terms that describe the image contents when retrieving similar images. We plan to enhance the current framework by considering semantic term similarities extracted from multiple biomedical ontologies and complementary quantitative imaging descriptors. In the future, we also plan to involve this system into larger clinical studies.

This project was funded by grants from NCI, NIH (# U01CA142555-01), SNSF (# PBGEP2_142283), and GE Medical Systems.

5. REFERENCES

[1] G. D. Rubin, "Data explosion: The challenge of multidetector-row CT," *European Journal of Radiology*, vol. 36, no. 2, pp. 74–80, 2000.

[2] P. Aigrain, H. Zhang, and D. Petkovic, "Content-based representation and retrieval of visual media: A state-of-the-art review," *Multimedia Tools and Applications*, vol. 3, pp. 179–202, 1996.

[3] J. C. Van Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek, "Visual word ambiguity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1271–1283, 2010.

[4] A. Mojsilovic and B. Rogowitz, "Capturing image semantics with low-level descriptors," in *Proceedings of the IEEE International Conference on Image Processing*, 2001, vol. 1, pp. 18–21.

[5] J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-fei, "ImageNet: A large-scale hierarchical image database," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[6] S. A. Napel, C. F. Beaulieu, C. Rodriguez, J. Cui, J. Xu, A. Gupta, D. Korenblum, H. Greenspan, Y. Ma, and D. L. Rubin, "Automated retrieval of CT images of liver lesions on the basis of image similarity: Method and preliminary results," *Radiology*, vol. 256, no. 1, pp. 243–252, 2010.

[7] H. Ma, J. Zhu, M. R. T. Lyu, and I. King, "Bridging the semantic gap between images and tags," *IEEE Transactions on Multimedia*, vol. 12, no. 5, pp. 462–473, 2010.

[8] D. Zhang, M. M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346–362, 2012.

[9] C. Kurtz, A. Depeursinge, S. Napel, C. F. Beaulieu, and D. L. Rubin, "On combining image-based and ontological semantic similarities for medical image retrieval applications," *Submitted to the Medical Image Analysis journal*, 2014.

[10] C. Kurtz, N. Passat, P. Gançarski, and A. Puissant, "A histogram semantic-based distance for multiresolution image classification," in *Proceedings of the IEEE International Conference on Image Processing*, 2012, pp. 1157–1160.

[11] C. Kurtz, P. Gançarski, N. Passat, and A. Puissant, "A hierarchical semantic-based distance for nominal histogram comparison," *Data & Knowledge Engineering*, vol. 87, no. 1, pp. 206–225, 2013.

[12] A. Depeursinge, A. Foncubierta-Rodriguez, D. Ville, and H. Müller, "Multiscale lung texture signature learning using the Riesz transform," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, N. Ayache, H. Delingette, P. Golland, and K. Mori, Eds., vol. 7512 of *LNCS*, pp. 517–524. Springer, 2012.

[13] A. Depeursinge, A. Foncubierta-Rodriguez, D. Van de Ville, and H. Müller, "Rotation-covariant texture learning using steerable Riesz wavelets," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 898–908, 2014.

[14] Curtis P Langlotz, "RadLex: A new method for indexing online educational materials," *RadioGraphics*, vol. 26, no. 6, pp. 1595–1597, 2006.

[15] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Transactions on Information Systems*, vol. 20, no. 4, pp. 422–446, 2002.

[16] Y. Rubner, C. Tomasi, and L. J. Guibas, "The Earth Mover's Distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, pp. 99–121, 2000.