Predicting Visual Semantic Descriptive Terms From Radiological Image Data: Preliminary Results With Liver Lesions in CT

Adrien Depeursinge*, Camille Kurtz, Christopher Beaulieu, Sandy Napel, and Daniel Rubin

***Abstract*—We describe a framework to model visual semantics of liver lesions in CT images in order to predict the visual semantic terms (VST) reported by radiologists in describing these lesions. Computational models of VST are learned from image data using linear combinations of high-order steerable Riesz wavelets and support vector machines (SVM). In a first step, these models are used to predict the presence of each semantic term that describes liver lesions. In a second step, the distances between all VST models are calculated to establish a nonhierarchical computationally-derived ontology of VST containing inter-term synonymy and complementarity. A preliminary evaluation of the proposed framework was carried out using 74 liver lesions annotated with a set of 18 VSTs from the RadLex ontology. A leave-one-patient-out cross-validation resulted in an average area under the ROC curve of 0.853 for predicting the presence of each VST. The proposed framework is expected to foster human–computer synergies for the interpretation of radiological images while using rotation-covariant computational models of VSTs to 1) quantify their local likelihood and 2) explicitly link them with pixel-based image content in the context of a given imaging domain.**

***Index Terms*—Computer-aided diagnosis (CAD), liver computed tomography (CT), RadLex, Riesz wavelets, steerability, visual semantic modeling.**

I. Introduction

MEDICAL imaging aims to support decision making by providing visual information about the human body. Imaging physics has evolved to assess the visual appearance of almost every organ with both high spatial and temporal resolution and even functional information. The technologies to preprocess, transmit, store, and display the images are implemented in all modern hospitals. However, clinicians rely nearly exclusively on their image perception skills for the final diagnosis [1]. The increasing variability of imaging protocols and the enormous amounts of medical image data produced per day in modern hospitals constitute a challenge for image interpretation, even for experienced radiologists [2]. As a result, errors and variations in interpretations are currently representing the weakest aspect of clinical imaging [3].

Successful interpretation of medical images relies on two distinct processes: 1) identifying important visual patterns and 2) establishing potential links among the imaging features, clinical context, and the likely diagnoses [4]. Whereas the latter requires a deep understanding and comprehensive knowledge of the radiological manifestations and clinical aspects of diseases, the former is closely related to visual perception [5]. A large-scale study on malpractice in radiology showed that the majority of errors in medical image interpretation are caused by perceptual misinterpretation [6]. Strategies for reducing perceptual errors includes the normalization of viewing conditions, sufficient training of the observers, availability of similar images and clinical data, multiple reporting, and image quantification [3]. The use of structured visual terminologies based on radiology semantics is also a promising approach to enable unequivocal definition of imaging signs [7], but is yet little used in routine practice.

Computerized assistance for image analysis and management is expected to provide solutions for the aforementioned strategies by yielding exhaustive, comprehensive, and reproducible data interpretation [8]. The skills of computers and radiologists are found to be very complementary, where high-level image interpretation based on computer-generated image quantification and its clinical context remains in the hands of the human observer [9]. However, several challenges remain unsolved and require further research for a successful integration of computer-aided diagnosis (CAD) into the radiology routine [10]. A fundamental question for a seamless workflow integration of CAD is to maximize interactions between CAD outputs and human conceptual processing [9]. The latter relies both on the trust and intuition of the user of the CAD system. Trust can only be achieved when a critical performance level is achieved by the system [10]. Intuition still requires extensive research efforts to design computer-generated outputs that match human semantics in radiology [7]. The transparency of the computer algorithms should be maximized so that the users can identify errors.

A. Semantic Information in Radiology Images

Radiologists rely on many visual terms to describe imaging signs relating to anatomy, visual features of abnormality and

Manuscript received March 19, 2014; revised April 25, 2014; accepted April 27, 2014. Date of publication May 01, 2014; date of current version July 30, 2014. This work was supported in part by the Swiss National Science Foundation (PBGEP2_142283), in part by the National Cancer Institute , and in part by the National Institutes of Health (U01-CA-142555 and R01–CA–160251). *Asterisk indicates corresponding author.*

*A. Depeursinge is with the Department of Radiology, School of Medicine, Stanford University, CA 94305 USA (e-mail: adepeurs@stanford.edu).

C. Kurtz is with the Department of Radiology of the School of Medicine, Stanford University, CA 94305 USA, also with the LIPADE (EA2517), University Paris Descartes, 75270 Paris, France.

C. Beaulieu, S. Napel, and D. Rubin are with the Department of Radiology, School of Medicine, Stanford University, CA 94305 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2014.2321347

diagnostic interpretation [11]. The vocabulary used to communicate these visual features can vary greatly [12], which limits both clear communication between experts and the formalization of the diagnostic thought process [13]. The use of standardized terminologies of imaging signs including their relations (i.e., ontologies) has been recently recommended to unambiguously describe the content of radiology images [3]. The use of biomedical ontologies in radiology opens avenues for enabling clinicians to access, query and analyze large amounts of image data using an explicit information model of image content [14].

There are several domain-specific standardized terminologies being developed, including the breast imaging reporting and data system (BI-RADS) [15], the Fleishner society glossary of terms for thoracic imaging [16], the nomenclature of lumbar disc pathology [17], the reporting terminology for brain arteriovenous malformations [18], the computed tomography (CT) colonography reporting and data system [19], the visually accessible Rembrandt images (VASARI) for describing the magnetic resonance (MR) features of human gliomas, as well as others [20]. Most of the standardized terminologies do not include relations among their terms (e.g., synonyms or hierarchical taxonomic structures), and as such, they thus do not constitute true ontologies. As a result, these terminologies cannot be used for semantic information processing leveraging interterm similarities. Recent efforts from the Radiological Society of North America (RSNA) were undertaken to create RadLex[1], an ontology unifying and regrouping several of the standardized terminologies mentioned above [21], in addition to providing terms unique to radiology that are missing from other exiting terminologies. RadLex contains more than 30 000 terms and their relations, which constitutes a very rich basis for reasoning about image features and their implications in various diseases.

B. Linking Image Contents to Visual Semantics

The importance of ontologies in radiology for knowledge representation is well established, but the importance of explicitly linking semantic terms with pixel-based image content has only recently been emphasized[2] [7]. Establishing such a link constitutes a next step to computational access to exploding amounts of medical image data [22]. This also provides the opportunity to assist radiologists in the identification and localization of diagnostically meaningful visual features in images. The creation of computational models of semantic terms may also allow the establishment of distances between the terms that can be learned from data, which can add knowledge about the meaning of semantic terms within existing ontologies.

C. Related Work

The bag-of-visual-words (BOVW) approach [23] aims at discovering visual terms in an unsupervised fashion to minimize the semantic gap between low-level image features alone and higher-level image understanding. The visual words (VW) are defined as the cluster centroids obtained from clustering the image instances expressed in a given low-level feature space. Its ability to enhance medical image classification and retrieval when compared to using the low-level features was demonstrated by several studies [24]–[28]. Attempts were made for visualizing the VWs, aiming at interpreting the visual semantics being modeled. In [24] and [26], color image overlays are used to mark the local presence of VWs in image examples. In [25], [27]–[29], prototype image patches (those closest to the respective VWs) are displayed to visualize the information modeled. Unfortunately, VWs often do not correspond to the actual semantics in medical images, and they are therefore very difficult to interpret for radiologists.

The link between VWs and medical VSTs is studied in [30], [31]. Liu *et al.* used supervised sparse auto-encoders to automatically derive several patterns (i.e., VWs) per disease. However, albeit the learned patterns were derived from the disease classes, they did not correspond to visual semantics belonging to a controlled vocabulary and therefore did not have a clear semantic interpretation [31]. In the context of endomicroscopic video retrieval, André *et al.* use a Fisher-based approach to learn the links between VWs learned from dense-scale–invariant feature transform (SIFT) and eight visual semantic terms (VST) [30]. Entire videos can be summarized by star plots reflecting the presence of VSTs, but the transparency of the algorithms remains limited as the occurrence of the VSTs are not localized in the images.

Other studies have focused on the direct modeling of VSTs from application-specific semantic vocabularies [32]–[37]. In the context of histological image retrieval, Tang *et al.* built semantic label maps that localize the occurrence of VSTs from Gabor and color histogram features [33]. The authors further refined their VST maps using spatial-hidden Markov models in [38]. In [34], the link between ad-hoc low-level image features based on gray-level intensity and VSTs can be tailored for each specific user for describing high-resolution CT (HRCT) images of the lungs. Shyu *et al.* evaluated the discriminatory power of low-level computational features (i.e., gray-level intensity) for predicting human perceptual categories in HRCT in[39]. In [40], a comprehensive set of low-level image features (i.e., shape, size, gray-level intensity, and texture) was used to probabilistically model lung nodule image semantics. Kwitt *et al.* [36] defined semantic spaces by assuming that VSTs in endoscopic images are living on Riemannian manifolds in a space spanned by SIFT features. They could derive a positive-definite semantic kernel that can be used with support vector machine (SVM) classifiers. Gimenez *et al.* [37] used a comprehensive feature set including state-of-the-art contrast, texture, edge, and shape features together with LASSO (least absolute shrinkage and selection operator [41]) regression models to predict the presence of VST from entire ROIs.

The above-mentioned studies demonstrated the feasibility of predicting VSTs from lower-level image features. However, both transparency and performance of most systems suffer from two limitations. First, the automatic annotation of global regions of interest (ROI) can be ambiguous [32]. Local quantifications of the VSTs can increase the transparency of the system by highlighting visual features that are recognized as positive inside the ROI. Second, most of the studies do not allow for

[1]Online: https://www.rsna.org/RadLex.aspx , as of March 2014.

[2]Liver annotation Task at ImageCLEF, http://www.imageclef.org/2014/liver/ , as of March 2014.

rotation-invariant detection of the VSTs by relying on low-level computational features that are analyzing images along arbitrary directions (e.g., SIFT, oriented filterbanks, gray-level co-occurrence matrices). VSTs are typically characterized by directional information (e.g., lesion boundary, nodule, vascular structure), but their local orientation may vary greatly over the ROI. Optimal modeling of VSTs requires image operators that are rotation-covariant, enabling the modeling of the local relative organization of the directions independently from the orientation of the VST [42]. The importance of the local relative orientation of directions for classifying normal liver tissue versus cancer tissue has also been highlighted by Upadhyay *et al.* using 3-D rigid motion invariant texture features [43].

In this work, we learn rotation-covariant computational models of RadLex VSTs from the visual appearance of liver lesions in CT. The models are built from linear combinations of Nth-order steerable Riesz wavelets, which are learned using SVMs (see Section II-C). This allows local alignment of the models to maximize their response, which can be computed analytically for any order N. The scientific contribution of the VST models is twofold. First, the models can be used to predict and quantify the local likelihood of VSTs in ROIs. The latter is computed as the dot product between 12×12 image patch instances and one-versus-all (OVA) SVM models in a feature space spanned by the energies of the magnitudes of locally-steered VST models. Second, Euclidean distances are computed for every pair of VST models to establish a nonhierarchical computationally-derived ontology containing inter-term synonymy and complementarity. This work constitutes, to the best of our knowledge, a first attempt to establish a direct link between image contents and the visual semantics used by radiologists to interpret images.

TABLE I
VSTs FROM RADLEX USED TO DESCRIBE THE APPEARANCE OF THE LIVER LESIONS IN CT SCANS. THE 18 VSTs DESCRIBING THE MARGIN AND THE INTERNAL TEXTURE OF THE LESIONS ARE MARKED IN BOLD

category	VST	frequency	patch location
lesion margin	**1) circumscribed margin**	70.3 %	peripheral
	2) irregular margin	12.2 %	
	3) lobulated margin	12.2 %	
	4) poorly–defined margin	16.2 %	
	5) smooth margin	45.9 %	
lesion substance	**6) internal nodules**	12.2 %	internal
perilesional tissue characterization	**7) normal perilesional tissue**	43.2 %	peripheral
lesion focality	8) solitary lesion	37.8 %	—
	9) multiple lesions 2-5	21.6 %	
	10) multiple lesions 6-10	20.3 %	
	11) multiple lesions >10	18.9 %	
lesion attenuation	**12) hypodense**	72.2 %	internal
	13) soft tissue density	16.2 %	
	14) water density	14.9 %	
overall lesion enhancement	**15) enhancing**	62.2 %	peripheral
	16) hypervascular	14.9 %	internal
	17) nonenhancing	29.7 %	peripheral
spatial pattern of enhancement	**18) heterogeneous enh.**	13.5 %	internal
	19) homogeneous enh.	32.4 %	internal
	20) peripheral discont. nodular enh.	17.6 %	peripheral
temporal enhancement	21) centripetal fill–in	17.6 %	—
	22) homogeneous retention	18.9 %	
	23) homogeneous fade	21.6 %	
lesion uniformity	**24) heterogeneous**	41.9 %	internal
	25) homogeneous	56.8 %	
overall lesion shape	26) round	25.7 %	—
	27) ovoid	45.9 %	
	28) lobular	25.7 %	
	29) irregularly shaped	12.2 %	
lesion effect on liver	31) abuts capsule of liver	17.6 %	—

II. METHODS

Notations

A VST is denoted as c_i, with $i = 1, \ldots, I$ while its likelihood of appearance in an image f is denoted $a_i \in [0, 1]$.

A generic d-dimensional signal f indexed by the continuous-domain space variable $\boldsymbol{x} = \{x_1, x_2, \ldots, x_d\} \in \mathbb{R}^d$ is considered. The d-dimensional Fourier transform of f is noted as

$$f(\boldsymbol{x}) \overset{\mathcal{F}}{\longleftrightarrow} \hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\boldsymbol{x}) e^{-j\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle} \mathrm{d}x_1 \ldots \mathrm{d}x_d$$

with $\boldsymbol{\omega} = \{\omega_1, \omega_2, \ldots, \omega_d\} \in \mathbb{R}^d$.

A. Dataset: VSTs of Liver Lesions in CT

The institutional review board approved the retrospective analysis of de-identified patient images. The dataset consisted of 74 contrast-enhanced CT images of liver lesions in the portal venous phase with a slice thickness of 5 mm [22], [37]. There are eight diagnoses of the lesions: metastasis (24), cyst (21), hemangioma (13), hepatocellular carcinoma (6), focal nodular hyperplasia (5), abscess (3), laceration (1), and fat deposition (1). A radiologist (C.F.B., 15 years of abdominal CT experience) used the axial slice f with the largest lesion area to circumscribe the liver lesions, producing image ROIs. Each lesion was annotated with an initial set of 72 VSTs from the RadLex ontology [22]. The presence of a term c_i did not imply the absence of all others, where each lesion can contain multiple VSTs. All terms with an appearance frequency[3] below 10% and above 90% were discarded, resulting in an intermediate set of 31 terms. A final set of 18 VSTs describing the margin and the internal texture of the lesions was used, which excludes terms describing the overall shape of the lesion (see Table I). Each of the 74 ROIs was divided into 12×12 patches[4] to analyze 1) the margin of the lesion (i.e., periphery) and 2) the internal texture of the lesion. The distinction between these zones of a lesion is relevant to understanding how the algorithm performs, and also parallels how radiologists interpret lesions, taking into consideration both the boundary and internal features. The peripheral patches 1) were constrained to have their center on the ROI boundary and the internal patches, 2) had to have their four corners inside the ROIs (see Fig. 1). The patches were overlapping with a minimum distance between the centers equals to one pixel. A maximum of 100 patches were randomly selected per ROI (i.e., 50 peripheral and 50 internal). Each of the 18 VST were represented by every patch extracted from the corresponding ROIs. Peripheral or internal patches were used as image instances depending on the VST's localizations

[3]The appearance frequency of a VST is defined as the percentage of lesions in the database in which the term was present.

[4]Patches larger than 12×12 did not fit in the smallest lesion.

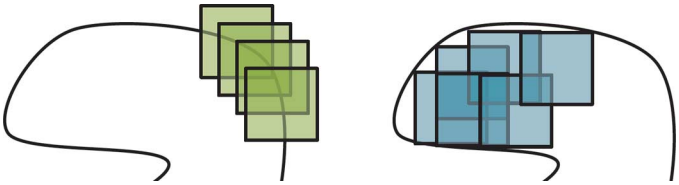

Fig. 1. Location of the image patches for the characterization of the margin (left) and the internal texture (right) of the lesions.

(see Table I). Peripheral patches were used for *enhancing* and *nonenhancing* to model the transition of enhancement at the boundary of the lesion in the portal venous phase.

B. Rotation-Covariant VST Modeling

Recent work showed that the Riesz transform and its multi-scale extension constitutes a very efficient computational model of visual perception [44], since it performs multi-directional and multi-scale image analysis while fully covering the angular and spatial spectrums. This constitutes a major advantage when compared to other approaches relying on arbitrary choices of scales or directions for analysis [i.e., Gabor wavelets, gray-level co-occurrence matrices (GLCM), local binary patterns (LBP)]. In a first step, we create VST models using linear combinations of Nth-order steerable Riesz wavelets. Then, the VST models are steered locally to maximize their response, which can be done analytically for any order N and yields rotation-covariant results [42].

1) Steerable Riesz Wavelets: Steerable multi-directional and multi-scale image analysis is obtained using the Riesz transform. Steerable Riesz wavelets are derived by coupling the Riesz transform and an isotropic multi-resolution framework[5][45]. The Nth-order Riesz transform $\boldsymbol{\mathcal{R}}^N$ of a 2–D function f yields $N+1$ components $\mathcal{R}^{(n,N-n)}$, $n=0,1,\ldots,N$ that form multi-directional filterbanks. Every component $\mathcal{R}^{(n,N-n)} \in \boldsymbol{\mathcal{R}}^N$ is defined in the Fourier domain as

$$\widehat{\mathcal{R}^{(n,N-n)}}\{f\}(\boldsymbol{\omega}) = \sqrt{\frac{N}{n!(N-n)!}} \frac{(-j\omega_1)^n(-j\omega_2)^{N-n}}{\|\boldsymbol{\omega}\|^N}\hat{f}(\boldsymbol{\omega}) \quad (1)$$

with $\omega_{1,2}$ corresponding to the frequencies along the two image axes $x_{1,2}$. The multiplication with $j\omega_{1,2}$ in the numerator corresponds to partial derivatives of f and the division by the norm of $\boldsymbol{\omega}$ in the denominator ensures that only phase information (i.e., directionality) is retained. The directions of every component is defined by Nth-order partial derivatives in (1).

The Riesz filterbanks are steerable, which means that the response of every component $\mathcal{R}^{(n,N-n)}$ oriented with an angle θ can by synthesized from a linear combination of all components $\mathcal{R}$ using a steering matrix $\boldsymbol{A}^\theta$ as in [42]

$$\boldsymbol{\mathcal{R}}^N\{f^\theta\}(\mathbf{0}) = \boldsymbol{A}^\theta \boldsymbol{\mathcal{R}}^N\{f\}(\mathbf{0}). \quad (2)$$

$\boldsymbol{A}^\theta$ contains the respective coefficients of each component $\mathcal{R}^{(n,N-n)}$ to be oriented with an angle θ.

[5]Simoncelli's multi-resolution framework is used with a dyadic scale progression. The scaling function is not used.

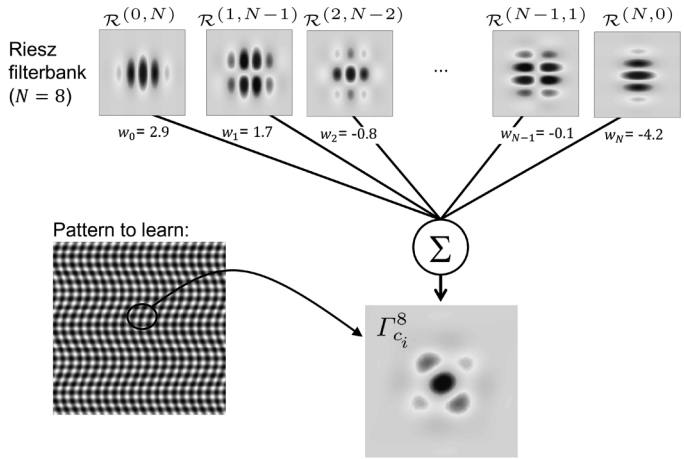

Fig. 2. Example of the construction of a model $\Gamma^8_{c_i}$ using a linear combination of the Riesz templates $\mathcal{R}^{(n,N-n)}$. $\Gamma^8_{c_i}$ is visually similar to the pattern [42].

2) Steerable VST Models: VST models are built using linear combinations of multi-scale Riesz components. Such models characterize the organizations of directions at various scales that are specific to each VST. At a fixed scale, a VST model $\Gamma^N_{c_i}$ is defined as

$$\begin{aligned}\Gamma^N_{c_i} &= \boldsymbol{w}^{\mathrm{T}}_{c_i}\boldsymbol{\mathcal{R}}^N \\ &= w_1\mathcal{R}^{(0,N)} + w_2\mathcal{R}^{(1,N-1)} + \cdots + w_{N+1}\mathcal{R}^{(N,0)}\end{aligned} \quad (3)$$

where $\boldsymbol{w}_{c_i}$ contains the weights of the respective Riesz components for the VST c_i. An example of the construction of a model for a synthetic pattern is shown in Fig. 2. By combining (2) and (3), the response of a model $\Gamma^N_{c_i}$ oriented by θ can still be expressed as a linear combination of the initial Riesz components as

$$\Gamma^{N,\theta}_{c_i} = \boldsymbol{w}^{\mathrm{T}}_{c_i}\boldsymbol{A}^\theta\boldsymbol{\mathcal{R}}^N. \quad (4)$$

l_2-norm SVMs are used to learn the optimal weights in a feature space spanned by the energies of concatenated multi-scale Riesz components to be optimally discriminant in OVA classification configurations [42], [46].

C. From VST Models to a Computationally-Derived Ontology

Once multi-scale models $\Gamma^N_{c_i}$ are learned for every VST using OVA configurations, the distance between every pair of VST can be computed as the Euclidean distance between the corresponding set of weights $\boldsymbol{w}_{c_i}$. The symmetric matrix $\Phi(c_i,c_j)$ can be considered as a nonhierarchical computationally-derived ontology modeling the visual inter-term synonymy and complementarity relations.

D. Rotation-Covariant Local Quantification of VSTs

The orientation θ of each model $\Gamma^N_{c_i,s_j}$ is optimized at each position $\boldsymbol{x}_p$ and for each scale s_j to maximize its local magnitude as

$$\theta_{\mathrm{dom},s_j}(\boldsymbol{x}_p) := \underset{\theta\in[0,\pi]}{\arg\max}\left(|\boldsymbol{w}^{\mathrm{T}}_{c_i,s_j}\boldsymbol{A}^\theta\boldsymbol{\mathcal{R}}^N_{s_j}\{f\}|\right)(\boldsymbol{x}_p). \quad (5)$$

The maximum magnitude of the model $\Gamma^N_{c_i,s_j}$ steered using $\theta_{\mathrm{dom},s_j}$ at the position $\boldsymbol{x}_p$ is computed as

$$m_{c_i,s_j}(\boldsymbol{x}_p) = \boldsymbol{w}^{\mathrm{T}}_{c_i,s_j}\boldsymbol{A}^{\theta_{\mathrm{dom},s_j}}\boldsymbol{\mathcal{R}}^N_{s_j}\{f\}(\boldsymbol{x}_p). \quad (6)$$

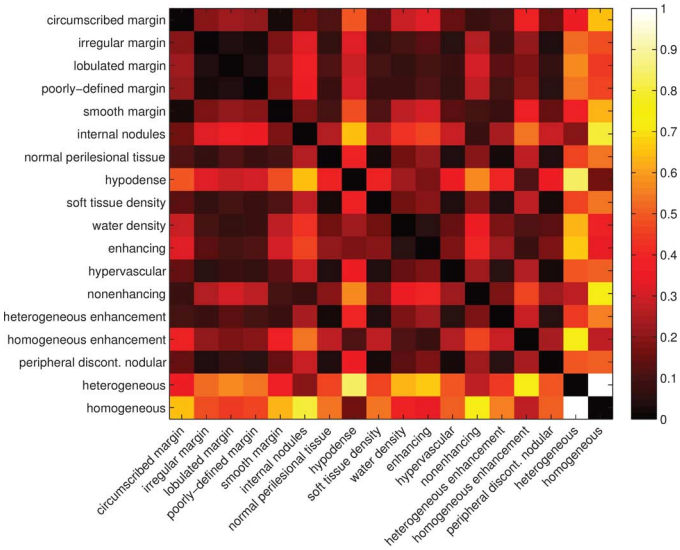

Fig. 3. Computationally-derived ontology matrix $\Phi(c_i, c_j)$ containing the normalized Euclidean distances for every pair of VST models. Values closer to zero (black) indicate the shortest distances, or most similar terms. Clear relationships between groups of VSTs are revealed both for terms describing lesion margin and internal texture.

For a given image patch, a feature vector $\boldsymbol{v}$ can be built as the energies E of the magnitudes over the patch of every steered model as

$$\boldsymbol{v} = \{E(m_{c_1,s_1}), \ldots, E(m_{c_1,s_J}), \ldots, E(m_{c_I,s_1}), \ldots, E(m_{c_I,s_J})\}. \quad (7)$$

The dimensionality of $\boldsymbol{v}$ is $I \times J$. It worth noting that the features from (7) are not steerable anymore after using the energies of the steered model's magnitudes. OVA SVM models $\boldsymbol{u}_i$ with Gaussian kernels $\phi(\boldsymbol{v})$ are used to learn the presence of every VST in the feature space spanned by the vectors $\boldsymbol{v}$ in (7). The decision value of the SVM for the image patch $\boldsymbol{v}_p$ measures the likelihood a_i of a VST as

$$a_i = \langle \phi(\boldsymbol{v}), \phi(\boldsymbol{u_i}) \rangle + b = \exp\left(\frac{-\|\boldsymbol{v} - \boldsymbol{u_i}\|^2}{2\sigma^2}\right) + b \quad (8)$$

where b is the bias of the SVM model. Likelihood maps are created by displaying a_i values from all overlapping patches.

E. Experimental Setup

The number of scales was chosen as $J = \lfloor \log_2(12) \rfloor = 3$ to cover the full spatial spectrum of 12×12 patches. The order of the Riesz transform $N = 8$ was used, which we found to provide an excellent trade-off between computational complexity and the degrees of freedom of the filterbanks in [42], [47]. A leave-one-patient-out (LOPO) cross-validation (CV) was used both to learn the VST models and to estimate the performance of VST detection using OVA configurations. For each fold, the training set was used both for learning the models and to train the SVMs in the feature space spanned by the vectors $\boldsymbol{v}$ in (7). No multi-class classification is performed since the VSTs are not considered as mutually exclusive [37]. The random selection of the patches was repeated five times to assess the robustness of the approach. The decision values a_i [see (8)] of the test patches were averaged over each ROI and used to build receiver operating characteristic (ROC) curves for each VST. The values a_i of the test patches were also used to locally quantify the presence of a VST and were color-coded to create VST-wise likelihood maps. CT intensities were used as additional features to model gray-level distributions in the patches, which we found to be complementary to Riesz models in [46]. The distribution of CT intensities in $[-60, 220]$ Hounsfield Units (HU) were divided into 20 histogram bins that were directly concatenated with the features obtained from (7).

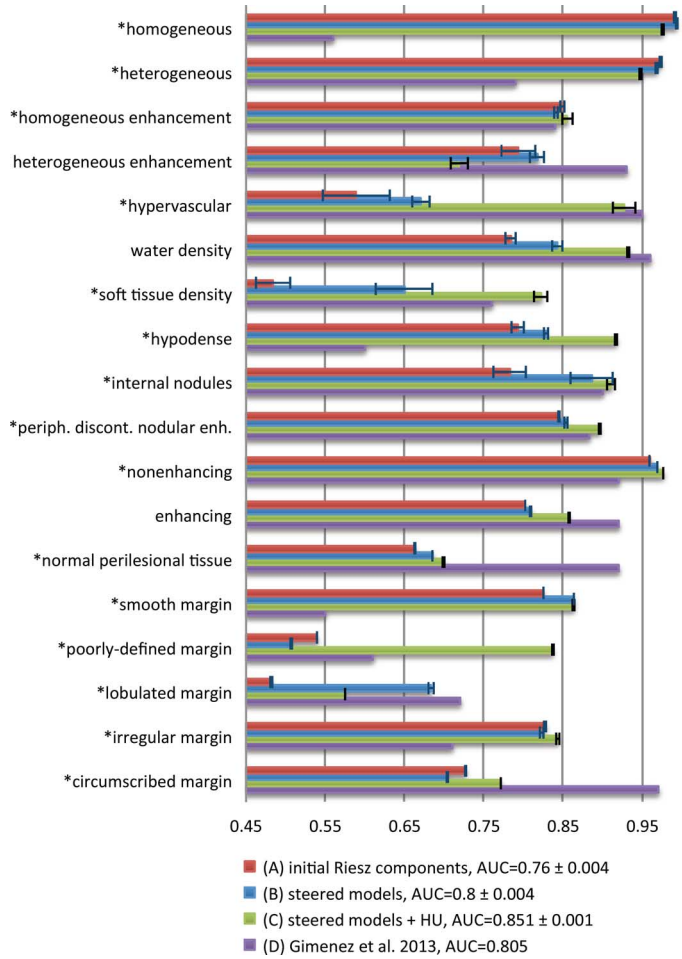

Fig. 4. Comparison of the automated detection performance between (A) initial Riesz components R^N, (B) steered models $\Gamma^N_{c_i}$, (C) steered models $\Gamma^N_{c_j}$ combined with HU histogram bins, and (D) the best results obtained by Gimenez *et al.* [37] on the same dataset in terms of AUCs. (*) denotes p-values below 0.05 for the comparison between (A) and the (B). (B) are always higher or close to the best performance of the (A), which highlights the importance of rotation-covariance. The difference between the global AUCs of (A) and (B) (i.e., 0.76 versus 0.8) is associated with a p-value of 7.5446e-152. (C) and (D) were shown to be complementary.

III. Results

A. VST Models and Computationally-Derived Ontology

Examples of scale-wise models $\Gamma^N_{c_i,s_j}$ are shown in Fig. 5 for six VSTs. The distributions of the weights w_{c_i,s_j} for every $N+1$ Riesz component $\mathcal{R}^{(0,N)}_{s_j}, \ldots, \mathcal{R}^{(N,0)}_{s_j}$ are shown with bar plots. The normalized computationally-derived ontology matrix $\Phi(c_i, c_j)$ containing the Euclidean distances between every pair of VST models is represented as a heatmap in Fig. 3.

B. Automated VST Detection and Quantification

ROC curves were built by varying thresholds on the decision values a_i. Fig. 4 compares the detection performance using the

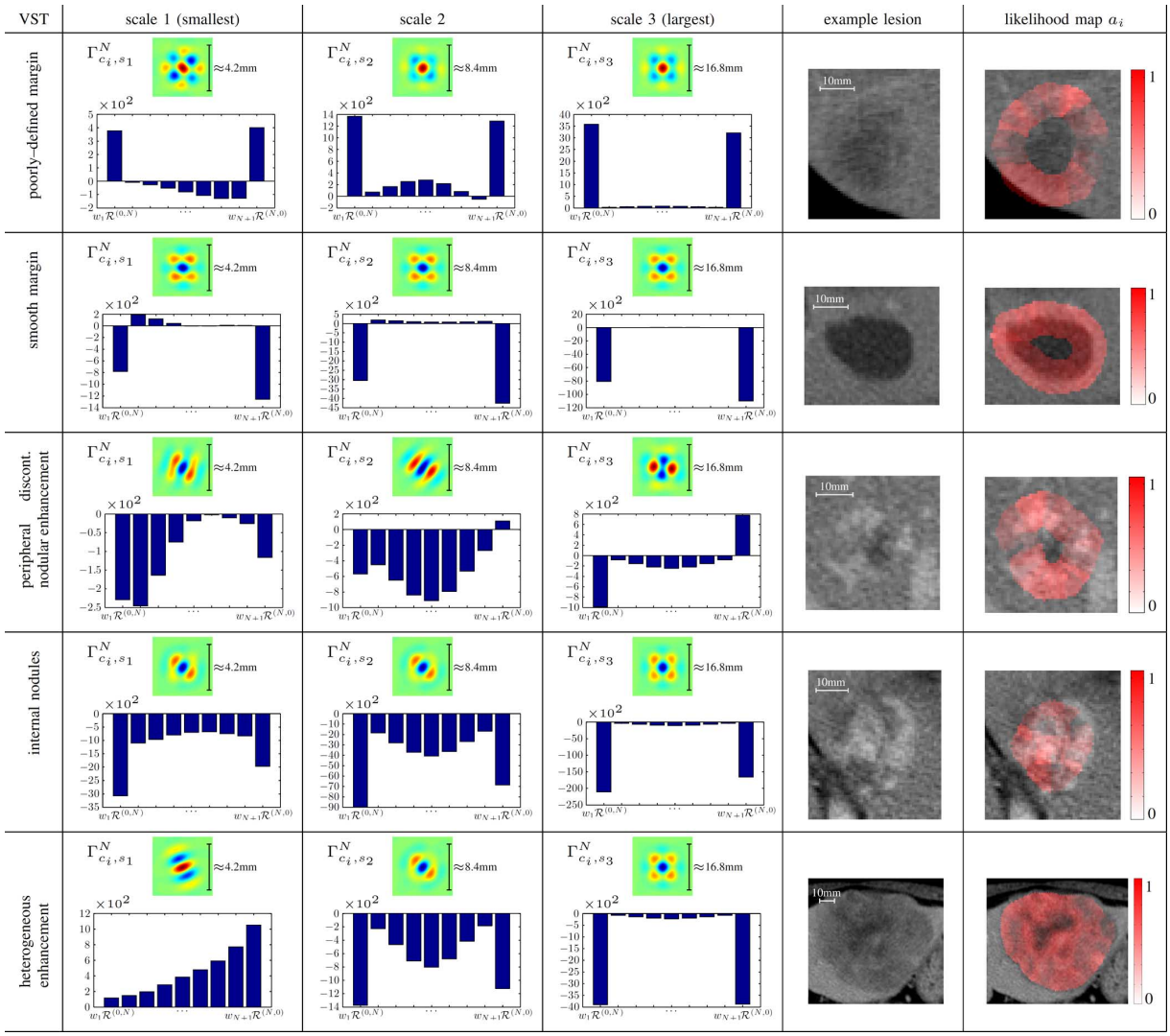

Fig. 5. Scale-wise models $\Gamma^N_{c_i,j}$ for five VSTs and $N = 8$. Distributions of the weights $w_{c_i,j}$ learned by OVA SVMs are represented for each scale with bar plots. Red regions in the last column of the table are showing examples of likelihood maps a_i computed from (8) for every VSTs.

energies of the magnitudes of steered models [i.e., (7)] versus the energies of the initial Riesz components. A paired T-test on a_i values is used to compare the two approaches. We also compared the performance of steered models combined with HU histogram bins with the best results obtained by Gimenez *et al.* on the same dataset as used here [37]. Examples of local quantification of the local presence a_i of VSTs (i.e., likelihood maps) are shown in the last column of Fig. 5.

IV. Discussions and Conclusion

We built computational models of human perceptual semantics in radiology. The framework identifies relevant organizations of image scales and directions related to VSTs in a rotation-covariant fashion using steerable Riesz wavelets and SVMs. The models were used both to (1) detect and quantify the local presence of VST and (2) to create a computationally-derived ontology from actual image content. In these early applications, our techniques for detection and quantification of VST enable automatic prediction of relevant semantic terms for radiologist consideration. This is an important step towards improving the accuracy of lesion description and diagnosis, with the aim of reducing overall interpretation errors. Our approach generates VST likelihood maps (see Fig. 5), which provide insights on the information modeled by the system, making it possible for the user to evaluate the amount of trust that can be put in its outputs and maximizes the transparency of the methods. Avoiding a "black box" type approach, this feature of our system is likely to be more intuitive to the ultimate end users. The automated detection and quantification of visual semantics grants access to large amounts of similar images by enabling interoperability with semantic indexing [22], [48]. The creation of a computationally-derived ontology from VST models can be used to refine and complement existing ontologies, where relations between terms are solely encoded by their hierarchical linguistic organization. The proposed computationally-derived ontology allows measuring inter-term synonymy and complementarity in the context of a given medical application adding additional useful knowledge to existing ontologies [49].

The visualization of the computationally-derived ontology matrix $\Phi(c_i, c_j)$ in Fig. 3 reveals clear relationships between groups of VSTs. Two homogeneous groups are found to model antonymous terms concerning the lesion margin: well- versus poorly-defined margin. *Irregular*, *lobulated*, and *poorly-defined margin* are very close, and they are all distant from *circumscribed* and *smooth margin*. The lack of distinction between the shape of the lesion and the type of lesion margin shows the inability of the proposed models to characterize overall lesion shape. For instance, *lobulated margin* and *poorly-defined margin* are not expected to be close, as it is possible to have lesions with margins that are both lobulated and circumscribed. Shared synonymy and antonymy is also observed in VSTs describing the internal texture of the lesions. *Heterogeneous* and *internal nodules* are found to share synonymy, and they are both opposed to *homogeneous*, *homogeneous enhancement*, and *hypodense*. It can also be observed that *hypervascular* is close to *heterogeneous enhancement* and *peripheral discontinuous nodular enhancement*, which all relate to the pattern of lesion enhancement. A few erroneous associations are observed (e.g., *soft tissue density* and *hypervascular*, or *water density* and *enhancing*). Overall, the computationally-derived ontology is found to be complementary to the RadLex ontology, because it allows connecting semantic concepts with their actual appearance in CT images. For instance, *heterogeneous* and *homogeneous* are very close to each other in RadLex because they both describe the uniformity of lesion enhancement, but they are opposed to each other in the computationally-derived ontology since they are visually antonymous in terms of texture characterization. The combination of the RadLex ontology and the computationally-derived ontology was shown to significantly improve image retrieval performance in [49]. Likewise, we would expect our computationally-derived ontology to be useful in combination with existing ontologies like RadLex in image retrieval and other applications.

Fig. 4 details the VST detection performance. It shows that although the steered models are not improving the results for all VSTs, they are always higher or close to the best performance of the initial Riesz components (i.e., global AUC equals to 0.8 versus 0.76, $p = 7.5446e\text{-}152$). An overall complementarity of the features based on steerable VST models and HU intensities is observed (e.g., *water density*, *soft tissue density*, *hypervascular*, *hypodense*). However, the detection performance is little improved or even harmed for texture-related terms when compared to using only steerable VST models (e.g., *heterogeneous*, *heterogeneous enhancement*, *homogeneous*). The proposed approach appears to be complementary to Gimenez *et al.* [37], where the errors are occurring for different VSTs. This suggests that the inclusion of other feature types can improve our approach.

Fig. 5 displays the relevance of the information modeled by a subset of VST models. The visualization of models $\Gamma^N_{c_i,j}$ reveals dominant scale-wise VSTs. The largest scale model of *peripheral discontinuous nodular enhancement* implements a detector of nodules surrounding the lesion boundary. The two smallest scale models of *internal nodules* are implementing circular dot detectors. These cases illustrate that the models of VSTs correspond to image features that actually describe the intended semantics. In general, the scale-wise distributions of the weights reveal the importance of Riesz components $\mathcal{R}^{(0,N)}$ and $\mathcal{R}^{(N,0)}$, which can be explained by their ability to model strongly directional structures (see Fig. 2). The stability of the models over the folds of the LOPO-CV is measured by the trace of the covariance matrices of VST-wise sets of weights w_{c_i} over the CV folds. The values of VST-wise traces are all below 0.33% of the trace of the global covariance matrix of all models from all folds. This demonstrates the stability, and hence the generalization ability, of the learned models.

We recognize several limitations of the current work, including a narrow imaging domain, a relatively small number of cases, and the use of somewhat thick, 5 mm CT sections. In future work, we plan to include additional cases to limit the risk of finding erroneous associations between terms caused by fortuitous co-occurrences of them. We plan to include additional image features that model the lesion shape. This will allow us to create separate computationally-derived ontologies based on the type of information modeled (i.e., intensity, texture, margin, and shape).

REFERENCES

[1] E. A. Krupinski, "The role of perception in imaging: Past and future," *Seminars Nucl. Med.*, vol. 41, no. 6, pp. 392–400, 2011.

[2] K. Andriole, J. Wolfe, and R. Khorasani, "Optimizing analysis, visualization and navigation of large image data sets: One 5000-section CT scan can ruin your whole day," *Radiology*, vol. 259, pp. 346–362, 2011.

[3] P. Robinson, "Radiology's Achilles' heel: Error and variation in the interpretation of the Rontgen image," *Br. J. Radiol.*, vol. 70, no. 839, pp. 1085–1098, 1997.

[4] G. Tourassi, S. Voisin, V. Paquit, and E. Krupinski, "Investigating the link between radiologists' gaze, diagnostic decision, and image content," *J. Am. Med. Inf. Assoc.*, pp. 1067–1075, 2013.

[5] B. Wood, "Visual expertise," *Radiology*, vol. 211, no. 1, pp. 1–3, 1999.

[6] L. Berlin, "Malpractice issues in radiology. Perceptual errors," *Am. J. Roentgenol.*, vol. 167, no. 3, pp. 587–590, 1996.

[7] D. L. Rubin, "Finding the meaning in images: Annotation and image markup," *Philos., Psychiatry, Psychol.*, vol. 18, no. 4, pp. 311–318, 2012.

[8] R. F. Wagner, M. F. Insana, D. G. Brown, B. S. Garra, and R. J. Jennings, *Texture Discrimination: Radiologist, Machine and Man in Vision*, C. Blakemore, K. Adler, and M. Pointon, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1991, pp. 310–318.

[9] E. A. Krupinski, "The future of image perception in radiology: Synergy between humans and computers," *Acad. Radiol.*, vol. 10, no. 1, pp. 1–3, 2003.

[10] B. v. Ginneken, C. M. Schaefer-Prokop, and M. Prokop, "Computer-aided diagnosis: How to move from the laboratory to the clinic," *Radiology*, vol. 261, no. 3, pp. 719–732, 2011.

[11] D. L. Rubin and S. Napel, "Imaging informatics: toward capturing and processing semantic information in radiology images," *Yearbook Med. Inf.*, pp. 34–42, 2010.

[12] J. L. Sobel, M. L. Pearson, K. Gross, K. A. Desmond, E. R. Harrison, L. V. Rubenstein, W. H. Rogers, and K. L. Kahn, "Information content and clarity of radiologists' reports for chest radiography," *Acad. Radiol.*, vol. 3, no. 9, pp. 709–717, 1996.

[13] E. S. Burnside, J. Davis, J. Chhatwal, O. Alagoz, M. J. Lindstrom, B. M. Geller, B. Littenberg, K. A. Shaffer, C. E. Kahn, and C. D. Page, "Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings," *Radiology*, vol. 251, no. 3, pp. 663–672, 2009.

[14] D. L. Rubin, P. Mongkolwat, V. Kleper, K. Supekar, and D. S. Channin, "Annotation and image markup: Accessing and interoperating with the semantic content in medical imaging," *IEEE Intell. Syst.*, vol. 24, no. 1, pp. 57–65, Jan./Feb. 2009.

[15] C. J. D'Orsi and M. S. Newell, "BI-RADS decoded: Detailed guidance on potentially confusing issues," *Radiol. Clin. N. Am.*, vol. 45, no. 5, pp. 751–763, 2007.

[16] D. M. Hansell, A. A. Bankier, H. MacMahon, T. C. McLoud, N. L. Müller, and J. Remy, "Fleischner society: Glossary of terms for thoracic imaging," *Radiology*, vol. 246, no. 3, pp. 697–722, 2008.

[17] D. F. Fardon, "Nomenclature and classification of lumbar disc pathology," *Spine*, vol. 26, no. 5, pp. 93–113, 2001.

[18] R. P. Atkinson *et al.*, "Reporting terminology for brain arteriovenous malformation clinical and radiographic features for use in clinical trials," *Stroke*, vol. 32, no. 6, pp. 1430–1442, 2001.

[19] M. Zalis, M. Barish, J. Choi, A. Dachman, H. Fenlon, J. Ferrucci, S. Glick, A. Laghi, M. Macari, E. McFarland, M. Morrin, P. Pickhardt, J. Soto, and J. Yee, "CT colonography reporting and data system: A consensus proposal," *Radiology*, vol. 236, no. 1, pp. 3–9, 2005.

[20] S. Goldberg, C. Grassi, J. Cardella, J. Charboneau, G. Iii, D. Dupuy, D. Gervais, A. Gillams, R. Kane, F. Jr, T. Livraghi, J. McGahan, D. Phillips, H. Rhim, S. Silverman, L. Solbiati, T. Vogl, B. Wood, S. Vedantham, and D. Sacks, "Image-guided tumor ablation: Standardization of terminology and reporting criteria," *J. Vascular Intervent. Radiol. Suppl.*, vol. 20, no. 7, pp. 377–390, 2009.

[21] D. L. Rubin, "Creating and curating a terminology for radiology: Ontology modeling and analysis," *J. Digital Imag.*, vol. 21, no. 4, pp. 355–362, 2008.

[22] S. 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.

[23] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Washington, DC, USA, 2003, vol. 2, pp. 1470–1477.

[24] B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, "Endomicroscopic video retrieval using mosaicing and visual words," in *Proc. IEEE Int. Symp. Biomed. Imag.: From Nano to Macro*, 2010, pp. 1419–1422.

[25] U. Avni, H. Greenspan, E. Konen, M. Sharon, and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 733–746, Mar. 2011.

[26] A. Foncubierta-Rodríguez, A. Depeursinge, and H. Müller, "Using multiscale visual words for lung texture classification and retrieval," in *Medical Content-Based Retrieval for Clinical Decision Support*. New York: Springer, 2012, vol. 7075, Lecture Notes Comput. Sci., pp. 69–79.

[27] A. Burner, R. Donner, M. Mayerhoefer, M. Holzer, F. Kainberger, and G. Langs, "Texture bags: Anomaly retrieval in medical images based on local 3D-texture similarity," in *Medical Content-Based Retrieval for Clinical Decision Support*, ser. Lecture Notes Comput. Sci.. New York: Springer, 2012, vol. 7075, pp. 116–127.

[28] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, "Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced CT images," *J. Digital Imag.*, vol. 25, no. 6, pp. 708–719, 2012.

[29] Y. Wang, T. Mei, S. Gong, and X.-S. Hua, "Combining global, regional and contextual features for automatic image annotation," *Pattern Recognit.*, vol. 42, no. 2, pp. 259–266, 2009.

[30] B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, "Learning semantic and visual similarity for endomicroscopy video retrieval," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1276–1288, Jun. 2012.

[31] L. Sidong, C. Weidong, S. Yang, P. Sonia, K. Ron, and F. Dagan, "A bag of semantic words model for medical content-based retrieval," in *MICCAI Workshop MCBR-CDS*. New York: Springer, 2013, LNCS, pp. 1–8.

[32] Y. Liu, F. Dellaert, and W. E. Rothfus, Classification driven semantic based medical image indexing and retrieval Robot. Inst., Carnegie Mellon Univ., Tech. Rep., 1998.

[33] L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam, "Extraction of semantic features of histological images for content-based retrieval of images," *SPIE Med. Imag. 1999, PACS Design Evaluat., Eng. Clin. Issues*, vol. 3662, pp. 360–368, 1999.

[34] A. S. Barb, C.-R. Shyu, and Y. P. Sethi, "Knowledge representation and sharing using visual semantic modeling for diagnostic medical image databases," *IEEE Trans. Inf. Technol. Biomed.*, vol. 9, no. 4, pp. 538–553, Dec. 2005.

[35] A. Mueen, R. Zainuddin, and M. S. Baba, "Automatic multilevel medical image annotation and retrieval," *J. Digital Imag.*, vol. 21, no. 3, pp. 290–295, 2008.

[36] R. Kwitt, N. Vasconcelos, N. Rasiwasia, A. Uhl, B. Davis, M. Häfner, and F. Wrba, "Endoscopic image analysis in semantic space," *Med. Image Anal.*, vol. 16, no. 7, pp. 1415–1422, 2012.

[37] F. Gimenez, J. Xu, Y. Liu, T. Liu, C. Beaulieu, D. L. Rubin, and S. Napel, "Automatic annotation of radiological observations in liver CT images," in *Proc. AMIA Annu. Symp.*, 2012, pp. 257–263.

[38] F. Yu and H. H. S. Ip, "Semantic content analysis and annotation of histological images," *Comput. Biol. Med.*, vol. 38, no. 6, pp. 635–649, 2008.

[39] C.-R. Shyu, C. Pavlopoulou, A. C. Kak, C. E. Brodley, and L. S. Broderick, "Using human perceptual categories for content-based retrieval from a medical image database," *Comput. Vis. Image Understand.*, vol. 88, no. 3, pp. 119–151, 2002.

[40] D. S. Raicu, E. Varutbangkul, J. D. Furst, and S. G. Armato III, "Modelling semantics from image data: opportunities from LIDC," *Int. J. Biomed. Eng. Technol.*, vol. 3, no. 1, pp. 83–113, 2010.

[41] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.

[42] A. Depeursinge, A. Foncubierta, D. Van De Ville, and H. Müller, "Rotation-covariant texture learning using steerable Riesz wavelets," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 898–908, Feb. 2014.

[43] S. Upadhyay, M. Papadakis, S. Jain, G. Gladish, I. A. Kakadiaris, and R. Azencott, "Semi-automatic discrimination of normal tissue and liver cancer lesions in contrast enhanced X-ray CT-scans," in *Abdominal Imaging. Computational and Clinical Applications*. Berlin, Germany: Springer, 2012, vol. 7601, pp. 158–167.

[44] K. Langley and S. J. Anderson, "The Riesz transform and simultaneous representations of phase, energy and orientation in spatial vision," *Vis. Res.*, vol. 50, no. 17, pp. 1748–1765, 2010.

[45] M. Unser and D. van De Ville, "Wavelet steerability and the higher-order Riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 636–652, Mar. 2010.

[46] A. Depeursinge, A. Foncubierta, D. van De Ville, and H. Müller, "Multiscale lung texture signature learning using the Riesz transform," in *Medical Image Computing and Computer-Assisted Intervention MICCAI*. Berlin, Germany: Springer, 2012, vol. 7512, pp. 517–524.

[47] A. Depeursinge, A. Foncubierta, H. Müller, and D. Van De Ville, "Rotation-covariant visual concept detection using steerable Riesz wavelets and bags of visual words," *SPIE Wavelets Sparsity XV*, vol. 8858, pp. 885811–885816, 2013.

[48] A. Gerstmair, P. Daumke, K. Simon, M. Langer, and E. Kotter, "Intelligent image retrieval based on radiology reports," *Eur. Radiol.*, vol. 22, no. 12, pp. 2750–2758, 2012.

[49] C. Kurtz, A. Depeursinge, S. Napel, C. F. Beaulieu, and D. L. Rubin, "On combining image-based and ontological dissimilarities for medical image retrieval applications," *Med. Image Anal.*, submitted for publication.