Clinical Experience Sharing by Similar Case Retrieval

Neda Barzegar Marvasti Ceyhun Burak Akgül Burak Acar
acarbu@boun.edu.tr
Bogazici University, Electrical and Electronics Department, Bebek, İstanbul, Turkey
Nadin Kökciyan Suzan Üsküdarlı Pınar Yolum
nadin.kokciyan@boun.edu.tr
Bogazici University, Computer Engineering Department, Bebek, İstanbul, Turkey
Rüştü Türkay Barış Bakır
rustuturkay@hotmail.com
İstanbul University, Faculty of Medicine, İstanbul, Turkey

ABSTRACT

Clinical experience sharing (CES) is a useful concept for both medical treatment and medical education purposes. One way of implementing CES is through the use of content based case retrieval (CBCR), where database of medical cases is browsed for case instances that are similar to the input query case. In this study, we introduce a new project called case retrieval in radiology (CaReRa), which aims at implementing CES for liver cases. We particularly focus on 3D liver images acquired by computed tomography (CT) and lay the foundations of a conceptual system outputting a ranked list of results for a given query case, formulated in this work as a liver lesion. A list of CT image features serves as computer generated descriptors together with user expressed annotations collected using a novel ontology of liver for radiology (ONLIRA). A two stage approach is proposed to utilize these two types of descriptors in cascade, namely semantic framing and similarity ranking. Initial retrieval performance results confirm the importance of ontology based descriptors, while also highlights the foci of future work needed to overcome the weaknesses.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

Clinical experience sharing; Content-based case retrieval; Semantic framing; Information storage and retrieval

MIIRH'13, October 22, 2013, Barcelona, Spain.

http://–enter the whole DOI string from rightsreview form confirmation.

1. INTRODUCTION

Accurate diagnostic procedures play an important role in medical field. Modern medical technologies have been offering some promising diagnostic techniques and machineries for medical cases, enabling expert medical doctors (MDs) to examine various aspects of the cases with higher diagnostic accuracy.

The experience of the medical professionals drastically affects the accuracy of the diagnoses, entertaining furthermore an additional motivation for MDs to share their clinical experiences and diagnostic routines. The concept of clinical experience sharing (CES) can be proposed as a searchable collective clinical experience platform, which enables practical, clinical, and educational knowledge sharing among large community of medical professionals. CES can also be leveraged to prevent situations, where lack of medical experience might have negative effects on diagnosis. In other words, by employing a CES platform, MDs can search for some specific cases and therefore retrieve the similar past cases from the collective database. This comparative procedure would be very helpful to give better and more careful diagnostic decisions. Moreover, such a platform can also be beneficial for educational purposes, where medical students can search for cases with similar characteristics but different diagnoses or vice versa.

A CES platform can be implemented by means of a content based case retrieval (CBCR) engine, searching for similar cases/patients in a database of past cases. Hence, there must be means of representing cases and some definitions for measuring similarity of different cases in the database. A case is composed of a wide variety of multi-modal information, such as patient demographics, medical history, lab test results, physical examination, genetics, drugs used, radiological images and image based observations/findings, etc. A CBCR engine is composed of both a content-based image retrieval (CBIR) engine employed in the visual (image) domain, and also a suite of similarity analysis tools in the non-image domain. The CBIR component utilizes both the image content and the other medical image metadata [1]. The image content is represented with a comprehensive set of computer generated features (CoG), while the image metadata is represented by user expressed features/annotations

(UsE) using a novel ontology (ONLIRA - Ontology of Liver for Radiology) [7]. By combining the non-image metadata as well as the knowledge models with CBIR, which focuses on finding similar images based on pixel analysis and image metadata, the retrieval system can be transformed from a conventional image-based search to patient/case-based reasoning [6],[3],[13],[2]. In [1], namely a CBCR.

It is important to note that building an all-inclusive CES system, which requires modelling the whole field of medicine, is currently not feasible. In much the same way as MDs themselves are specialized in discrete sub-domains of medicine and generally consider a limited part of medical information. Within the ongoing CaReRa (Case Retrieval In Radiology) project, we have chosen the liver as the application domain, mainly because of its importance and also relative ease of data collection. Our primary goal in CaReRa is to rank the liver cases in a liver case database according to their similarity to a query case.

In this paper, we focus on the CBIR component of the envisioned CBCR system, while discussing our approach to achieve the final goal of developing a CBCR for liver cases. More specifically, we propose the concept of *Semantic Framing* as an efficient way to combine the CoG features with UsE features. This paper is organized as follows: Section 2 introduces the CaReRa project. Section 2.1 reviews the CoG features, namely the image descriptors. In Section 2.2, we very briefly introduce ONLIRA and UsE features. Section 3 discusses the similarity analysis together with the *Semantic Framing*. Experiments are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes the paper.

2. THE CARERA PROJECT

The CaReRa project (*www.vavlab.ee.boun.edu.tr*), which aims at developing a CES space for liver cases, is the first step towards a multi-center consortium for further research and development over a broader range of medical domains. CaReRa is expected to greatly influence the medical society and attract major interest from the healthcare industry. The project includes 3D CT image processing (segmentation, characterization), image content analysis, ontology formation specific to the diagnosis of liver diseases, ontology guided query processing and case similarity/relevance analysis. Another important outcome of the project is the construction of a database including different liver cases, which are indeed annotated 3D CT images of liver augmented with multi-modal meta-data about the patient.

The overall objective is to retrieve old liver cases similar (or relevant with respect to the user's hidden goals) to a query case for the purposes of comparative diagnosis and medical education via sharing the collective experience over the medical community. Note that, in contrast to the more familiar CBIR concept, the query is a single patient with possibly missing information. These cases are described by metadata at 4 levels organized in a hierarchy as patient, study, series, and pathology. The patient-level includes demographics and medical history while the study-level contains the lab data and the clinical observations. The series-level contains the CT parameters and the liver (including the vessels) characteristics, while the pathology-level contains the lesion characteristics. The series-level and the pathology-level are represented with the CoG and the UsE features, where the UsE features are annotations associated with the novel ONLIRA ontology which is built upon the RadLex lexicon [11] with the extensions in attributes and additions of ontological relations between these attributes.

Table 1: The global image descriptors

Organ	Descriptor
Liver	Volume Mean Variance
Vessels	Percentage of vessel volume
Set of Lesions/Pathologies	Percentage of lesions Mean Variance Number of lesions Min volume of lesions Max volume of lesions

In general, there are two key issues at the forefront of the development of the CaReRa system. The first one is representing the case and the second one is to define the similarity between the cases. A liver case in CaReRa is represented with *i)* The computer generated image descriptors (CoG), *ii)* The image based user annotations (UsE), *iii)* The non-image based meta-data. These are explained in the following.

2.1 Computer Generated Image Descriptors (CoG)

The image descriptors capitalize on the visual cues contained in an image. In CaReRa, we consider image descriptors into two main domains based on their scope of application. The *global* image descriptors summarize the general visual properties of the liver, vessels and set of lesions. The *pathology* descriptors, on the other hand, reflect finer levels of visual information related to individual lesions.

The global image descriptors, listed in Table 1, cover the basic and liver-wide global statistical properties, such as mean, variance and volume. They are deduced directly from the CT volumes and the associated segmentation masks.

The pathology descriptors, listed in Table 2, are calculated for each single lesion in the liver separately. The descriptors marked with a $\times$ in Table 2 will be implemented in the upcoming phases of the CaReRa project. They are grouped in 5 types as geometric, locational, gray-scale, boundary and texture:

Geometric Descriptors can be deduced directly from the lesion mask and consist of shape-based cues such as surface area, sphericity, convexity, compactness, solidity, maximum-extent, and aspect-ratio.

Locational Descriptors refer to the location of the lesion with respect to the vessels and liver. The first feature, which is named proximity to vessels, indicates the lesion's distance to the nearest vessel. The touch area ratio with the vessels is the percentage of lesion surface area that touch a vessel. The liver segment identifier (ID) defines the lesion's location with respect to liver segmentation part such as locating in the right or the left lobe of liver.

Gray-scale Descriptors are derived directly from raw pixel intensities. These features include mean and variance of lesion tissue, asymmetry, kurtosis, portion of voxels having intensity higher than a defined

Table 2: The pathology descriptors and their parameters

Type	Name		
Geometric	Name	Cardinality	Status
	Volume	1	✓
	Surface area	1	✓
	Sphericity	1	✓
	Convexity	1	×
	Solidity	1	×
	Compactness	1	✓
	Maximum extent	1	✓
	Aspect ratio	1	✓
	Fourier descriptor	1	×
Locational	Proximity to vessel	1	✓
	Touch area ratio	1	✓
	Anatomical location	1	×
Boundary	Scale	30	✓
	Window	30	✓
GrayScale	Name	Cardinality	Status
	Mean	1	✓
	Variance	1	✓
	Thresholding	1	✓
	Histogram	64	✓
	Hist's abscissa	1	✓
	Hist's mean	1	✓
	Hist's variance	1	✓
	Hist's entropy	1	✓
	Smoothness	1	✓
	Assymmetry	1	✓
	kurtosis	1	✓
	Energy	1	✓
	Haar wavelet coeffs	8	✓
Texture	Name	Cardinality	Status
	Haralick	30	✓
	Tamura	5 hist	✓
	Gabor	52 hist	✓
	Hu moments	3	✓

threshold value, and gray-level histogram of every lesion. We also derive histogram-based features, namely, the bin of the histogram peak, the mean and the variance of histogram, the entropy of histogram, the low frequency coefficients of histogram's three level Haar wavelet transform [16] and the energy of histogram which is simply a measure of the uniformity of intensity in the histogram.

Boundary Descriptors capture the contrast and sharpness across the lesion boundaries as proposed in [19]. One dimensional gray-scale profiles are extracted along the normal direction of the lesion surface for ± 2 voxels . The intensity difference is the signed difference between the two ends of this profile. The sharpness refers to the rate of change of the CT values across the boundary. These two parameters are calculated via fitting a sigmoid function to the 1D boundary profile. The boundary feature vector for each lesion is composed of two $30 - bin$ histograms representing these two parameters over the whole lesion surface.

Texture Descriptors are part of photometric features and similar to gray-scale features, they are also derived from image intensity values. They encode spatial organization of pixel values of an image region and are divided into two groups: statistical model-based and transform-based. The *statistical model-based texture features* include Haralick's features and Tamura features. Six *Haralick's features* are derived from co-occurrence matrices in 5 distance and 13 directions [4]. In this work, the features have been calculated for distances of 1,2,4,6 and 8 in 13 degrees. For each distance, the final feature value was computed by averaging over the feature values corresponding to 13 angular directions. The Haralick features calculated from the co-occurrence matrix are angular second moment, entropy, inverse difference moment, interia, cluster prominence and cluster shade. Thus, a total of 30 texture characteristic was obtained, 6 characteristics for each voxel distance. The *Tamura features* correspond to human visual perception, which are coarseness, contrast, directionality, line-likeness, regularity and roughness. Generally, the first three features give us better description of the texture object [8]. Coarseness, which is the most fundamental texture feature, finds the largest size at which a texture exists. Contrast of the image is influenced by dynamic range of gray-levels in the image, polarization of the distribution of the black and white, sharpness of edges and period of repeating patterns, which show the picture quality [17]. Directionality calculates the direction of the texture along three axes. Hence,there are five image features including coarseness, contrast and 3 directionality images, which we store them as five feature vectors. The *transform based texture features* invoke standard transform domain analysis tools, the Gabor filter and the invariant Hu moments. We used Gabor filtering on the original CT image in 4 scales and 13 direction. As a result, 52 feature vectors are generated.[14]. The *Invariant Hu moments* are the set of moments, which are invariant under translation, scale and rotation [12]. We used the first three Hu moments in 3D domain.

2.2 Image Based User Annotations (UsE)

Radiologists describe their observations of CT scans in free-text radiology reports. Unfortunately, reports written in natural language are not suitable for automatic processing. Moreover, it becomes difficult to retrieve valuable information from reports because of their unstructured nature. Hence, it is beneficial to represent the observations in a structured manner. To represent imaging observations of the liver domain, ONLIRA has been developed. An ontology is a conceptualization of a domain in a structured way. It enables description of concepts and their properties as well as relationships between the concepts. ONLIRA consists of 40 concepts, 12 relationships and 36 properties. Currently, a data collection tool is using ONLIRA for the collection of imaging observations of the liver. ONLIRA is available for academic use at the CaReRa project website (*www.vavlab.ee.boun.edu.tr*).

2.3 Non-Image Based Meta-Data

The non-image based meta-data is grouped as patient (case) information and the study information. The data is composed of both general quantitative and qualitative prop-

erties as well as the one that are used commonly in the diagnosis of liver diseases. The utilization of this information is beyond the scope of this paper, hence a brief description is provided below.

The patient (case) information is described in terms of demographic information, regular use of drugs, chronic diseases and past surgeries. Demographics information includes age and sex of the patient. Description of the remaining metadata information is enhanced with standardized terminologies when possible. ATC codes [1] are used to specify regular drugs used by the patient. Past, chronic and genetic diseases are described using ICD-10 codes [2] . Surgery information for patients is provided using Turkish Ministry of Health codes [3] . All of the meta-data associated with the case is qualitative except for age.

The study information consists of both qualitative and quantitative properties. They include the complaints, kept as free text for future reference, the physical examination such as blood pressure, pulse, signs of portal hypertension and of liver dysfunction, the pre-diagnosis and diagnosis information is detailed using the ICD-10 codes, the non-regular drugs are specified by the use of ATC codes and finally the numeric results of 26 blood tests.

3. SIMILARITY ANALYSIS

Within the scope of this paper, the similarity among the liver lesions (pathologies) is considered. As described above, the pathologies are represented using the CoG and the UsE features. The CoG features are numeric features while the UsE features are discrete labels referring to our novel ontology (ONLIRA). Our similarity analysis approach is based on using a cascaded system of UsE based pre-filtering, which is termed as *Semantic Framing*, and CoG based ranking, which uses the Euclidean distance in the real valued feature vector space. Our primary goal in this study is to assess the feasibility of semantic framing, which can potentially provide an intuitive and flexible user interaction through semantic query processing. Currently, the semantic framing is limited to using the controlled vocabulary defined in ONLIRA, without any reference to the relations between the ONLIRA concepts. Hence, we used ONLIRA as a lexicon which is an extended version of Radlex for liver. No feature selection or weighting is applied on the CoG features either.

3.1 Semantic Framing

We have used semantic framing to narrow the domain of the query response using the method called formal concept analysis (FCA). A formal concept is a pair consisting of a set of objects and a set of attributes. FCA is a method used to derive a concept hierarchy given a dataset consisting of objects and their attributes [18]. The concept hierarchy can be defined as a concept lattice [15].

We use FCA as a conceptual knowledge representation of lesions. Basically, we have a set of lesions that consist of several properties, and we would like to find similar lesions. We define concepts by identifying lesions sharing the same values for a certain set of properties. Then, we use a lattice to represent the concept hierarchy of defined concepts. If a query lesion is given, the corresponding concept is identified in the lattice. Then, the query lesion is compared with every other lesion of the corresponding concept. If there are fewer lesions than a present number K (= 10), then the lesion query is expanded by selecting the upper neighbours of the corresponding concept. In the envisioned user interaction, K will not be fixed but rather the user will be allowed to navigate within the search space freely.

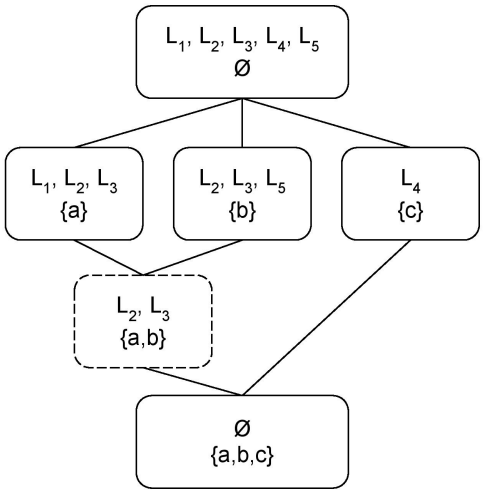

Figure 1: An example for lattice of lesions.

In Figure 1, a concept hierarchy of five lesions is depicted as a lattice. Rectangles represent formal concepts consisting of lesions (L_x) and their properties (a subset of $\{a, b, c\}$). If the query lesion is L_2, then dashed concept is identified in the lattice. In this case, there is only one lesion (L_3) to compare with. To get more results, the result set of lesions are enlarged with L_2's upper level lesions where the final set becomes $\{L_1, L_3, L_5\}$. L_2 is then compared to each lesion of the final set.

3.2 Similarity Ranking

We used the Euclidean distance for similarity ranking. Since all features, including both CoG and UsE features, have different value ranges, all are scaled and shifted to the range $[0, 1]$. While CoG features and some UsE features are real valued in $[0, 1]$, the remaining UsE features take discrete values from $\{0, 1\}$. The histograms are normalized by dividing every value in the histogram vector by sum of the values in the histogram vector. The scaled feature vectors are $\mathbf{v}_{CoG}$ and $\mathbf{v}_{UsE}$. The computed distances are scaled with the number of features in each set for normalization. Hence, we get the following distance measures between cases i and j,

$$d_{CoG}(i,j) = \frac{|\mathbf{v}^i_{CoG} - \mathbf{v}^j_{CoG}|_2}{\sqrt{N}} \quad N = \#(CoG) \tag{1}$$

$$d_{UsE}(i,j) = \frac{|\mathbf{v}^i_{UsE} - \mathbf{v}^j_{UsE}|_2}{\sqrt{M}} \quad M = \#(UsE) \tag{2}$$

We ranked the lesions preselected by semantic framing, according to their similarity to the query lesion using d_{CoG}. For comparison, we ranked all lesions in the dataset, skipping the semantic framing, using *i)* $d_{all} = \frac{1}{2}(d_{CoG} + d_{UsE})$, i.e. the mean distance, *ii)* d_{CoG} only.

[1] http://www.whocc.no/atc_ddd_index/
[2] http://www.who.int/classifications/icd/en/
[3] https://skrs3.sagliknet.saglik.gov.tr/

4. EXPERIMENTS

4.1 Dataset and Methodology

15 portal venous phase 3D CT images of the liver, from 15 patients (eight men, seven women; mean age,56 years; age range, $39-88$ years), including 30 of three types (13 cysts, 7 hemangiomas, and 10 metastases) are used in our analysis [10]. The image acquisition parameters were $120kVp$, $140-400mAs$, and $2.5-5.0-mm$ section thickness. From every type of lesions, lesions which are typically considered by radiologists are selected.

The CoG features are extracted automatically using custom interactive segmentation software. The UsE features are the ONLIRA based annotations provided by a board certified radiologist (R.T.). The similarity ground truth is the same as the one used in [10]. This ground truth was generated by two experienced radiologists based on a single 2D slice view of each lesion. Similarities were graded on a scale of 3 (3: Similar, 2: Somewhat similar, 1: Dissimilar). The mean gradings of the two radiologists are used.

Leave-one-out method is used for performance evaluation. Each of the 30 lesions is selected as a query and the remaining 29 lesion set is used as the search domain. The normalized discounted cumulative gain (NDCG) [5] and the well-known precision-recall (PR) [9] curves are computed for each query. The individual NDCG and PR curves are averaged to get the overall performance curves.

PR curves show the performance in retrieving the similar lesions from the search domain. The binary ground truth required for PR analysis is generated from the graded ground truth by assigning 1 (=similar) to the lesion pairs with graded ground truth larger than 1 and 0 (=dissimilar) otherwise. Precision and recall, for $k \in \{1, 2, ..., 10\}$ retrieved lesions, are defined as,

$$Precision(k) = \frac{\#(similar\_images\_retrieved)}{k} \quad (3)$$

$$Recall(k) = \frac{\#(similar\_images\_retrieved)}{\#(similar\_images)} \quad (4)$$

Note that, although both precision and recall range in $[0, 1]$, recall is upper bounded by $\frac{K}{\#(similar\_images)}$, where $K = 10$ in this study, hence maximum recall may not reach 1 for some of the queries. Precision-recall curves have a distinctive saw-tooth shape. In case the next lower ranked retrieved lesion is dissimilar to the query, while the current one is similar, the precision would decrease while recall remains the same. The standard way to cope with this issue is using interpolated precision. The interpolated precision at a certain recall level is defined as the highest precision found for any recall level greater than or equal to the current recall. The interpolated precision values are computed for $Recall = [0, 0.05, 0.1 \ldots, 0.95, 1.0]$.

NDCG is a standard technique used to measure the effectiveness of information retrieval algorithms, when graded truth is available, as represented by our three-point similarity scale. NDCG(k) ($k \in \{1, 2, ..., 10\}$) indicates the similarity value of the k retrieved lesions compared with their similarities to the query lesion on the scale of 0 to 1. It is done based on position of the retrieved lesions in the ranked list. For a given k, higher NDCG(k) means more lesions similar to the query image are ranked ahead of dissimilar ones. Note that, NDCG(k) equal to 1 implying perfect retrieval of k images. NDCG is defined as the ratio of discounted cumulative gain (DCG) over ideal discounted cumulative gain (IDCG) as follows,

$$NDCG(k) = \frac{DCG(k)}{IDCG(k)} \quad (5)$$

$$DCG(k) = R_1 + \sum_{i=2}^{k} \frac{R_i}{log_2 i} \quad (6)$$

$$IDCG(k) = \max(DCG(k)) \quad (7)$$

where R_i is the relevancy value in position i and k is the number of retrieved images.

4.2 Results

Three sets of experiments were conducted with the leave-one-out procedure:

* **Experiment 1:** A set of lesions is filtered using semantic framing with respect to the query, such that the minimum number of lesions (≥ 10) is selected. The selected set of lesions is ranked using d_{CoG}.

* **Experiment 2:** All lesions in the search domain are ranked using $d_{all} = \frac{1}{2}(d_{CoG} + d_{UsE})$ with respect to the query, and the top ranking 10 lesions are used for evaluation.

* **Experiment 3:** All lesions in the search domain are ranked using d_{CoG} only with respect to the query, and the top ranking 10 lesions are used for evaluation.

Figure 2 shows the average NDCG and PR curves for all experiments. Note that, since the total number of retrieved lesions is bounded by $K = 10$ in all experiments (though in Experiment 1 semantic framing preselects the minimum number of lesions ≥ 10), the recall values are upper bounded, hence may not reach 1.0. We have pursued this approach to make Experiments 2 and 3 comparable with Experiment 1 which uses semantic framing as a pre-filtering stage. In order to present this effect, we also included, in Figures 2.b-d, the number of queries that reached a given recall value together with the total number of queries that can potentially reach that recall value for $K = 10$ retrieved lesions.

Figure 3 shows the set of 10 retrieved lesions for a single query, in each one of the three experiments.

5. DISCUSSION

The experimental results clearly indicate the potential of using UsE features. The performance of using CoG and UsE features together, even with a simple Euclidean distance based similarity measure, has been observed to perform superior to both using CoG features alone and CoG and UsE features in a two stage scheme where the UsE features are utilized for semantic framing, as depicted in Figure 2.

The differences in PR curves, in Figures 2.c and 2.d, as well as the number of queries reaching a certain recall value in comparison with the actual number of lesions that could have reached that recall values for 10 retrieved lesions show that UsE features are important in recalling the similar lesions. The lower PR performance in Experiment 1 suggests that the current semantic framing fails to utilize the information content of UsE features for recall. In addition to that, a comparison of the NDCG plots in Figure 2.a indicates the importance of UsE in ranking the retrieved lesions.

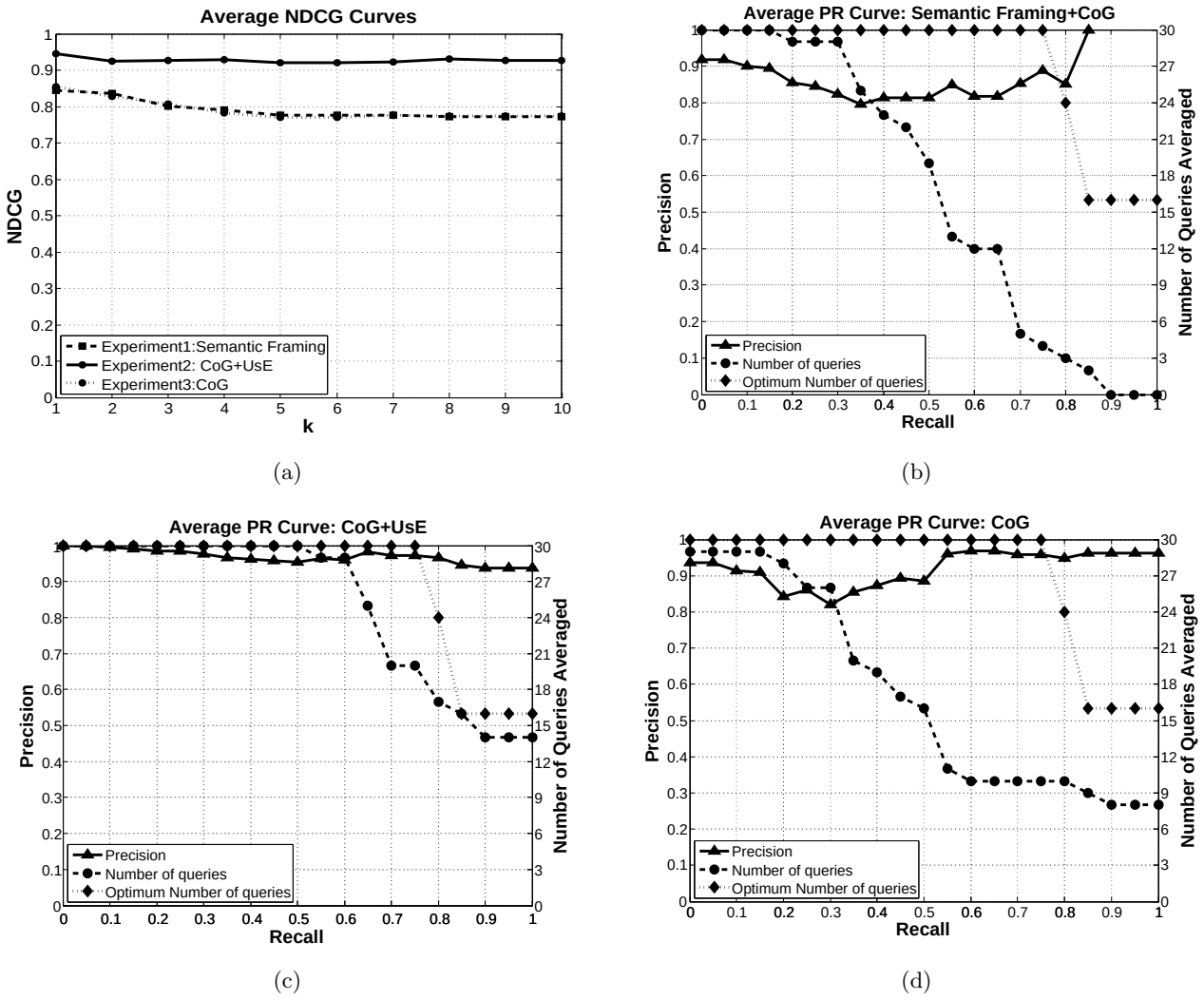

Figure 2: A comparison of retrieval performance results in all experiments. (a) The average NDCG plots as a function of the number of retrieved lesions, $k \in \{1, 2, ..., 10\}$. (b,c,d) The average PR curves, the number of queries that reach the computed recall values and the true number of queries that reach the same recall value. Only the top ranking 10 retrieved lesions are considered for Experiments 2 and 3.

On the other hand, the semantic framing provides a unique tool for intuitive query processing to navigate in the search space. It would facilitate search space *scaling* by means of query expansion/contraction as well as *panning* by means of manually editing the query's UsE features to focus on different subspaces of the search space. In a typical scenario where the user's intentions are unknown, such *scaling* and *panning* functionalities would be very beneficial to steer the searching towards the user's goals. Furthermore, the prefiltering by semantic framing would make more complex similarity analysis/ranking, such as manifold learning approaches (eg. Laplacian eigenmaps), feasible for large datasets. Currently, our semantic framing methodology does not utilize the whole information embedded in ONLIRA. Should the recall performance of semantic framing be improved, would the performance of the two stage scheme improve, remains to be investigated. The relations between the concepts in ONLIRA can be used during semantic framing to this effect.

6. CONCLUDING REMARKS

In this proof-of-concept study, we have laid preliminary foundations for CES and CBCR. We also have introduced the CaReRa project, focusing on a particular way of implementing CES for liver cases represented by 3D CT liver images and the novel ONLIRA ontology. We also introduced and employed a novel semantic framing method, which would be operational in interactive search space navigation by means of *scaling* and *panning* in a typical scenario where the users' intentions are unknown. We believe that such a navigation tool can help to overcome the drawbacks of classifier based approaches as they are optimized for predefined contexts, represented by the gold standards used, with respect to which the aforementioned classifiers are trained. However, despite these current results confirm the importance of ontology based features, the current simple semantic framing implementation failed to utilize this potential.

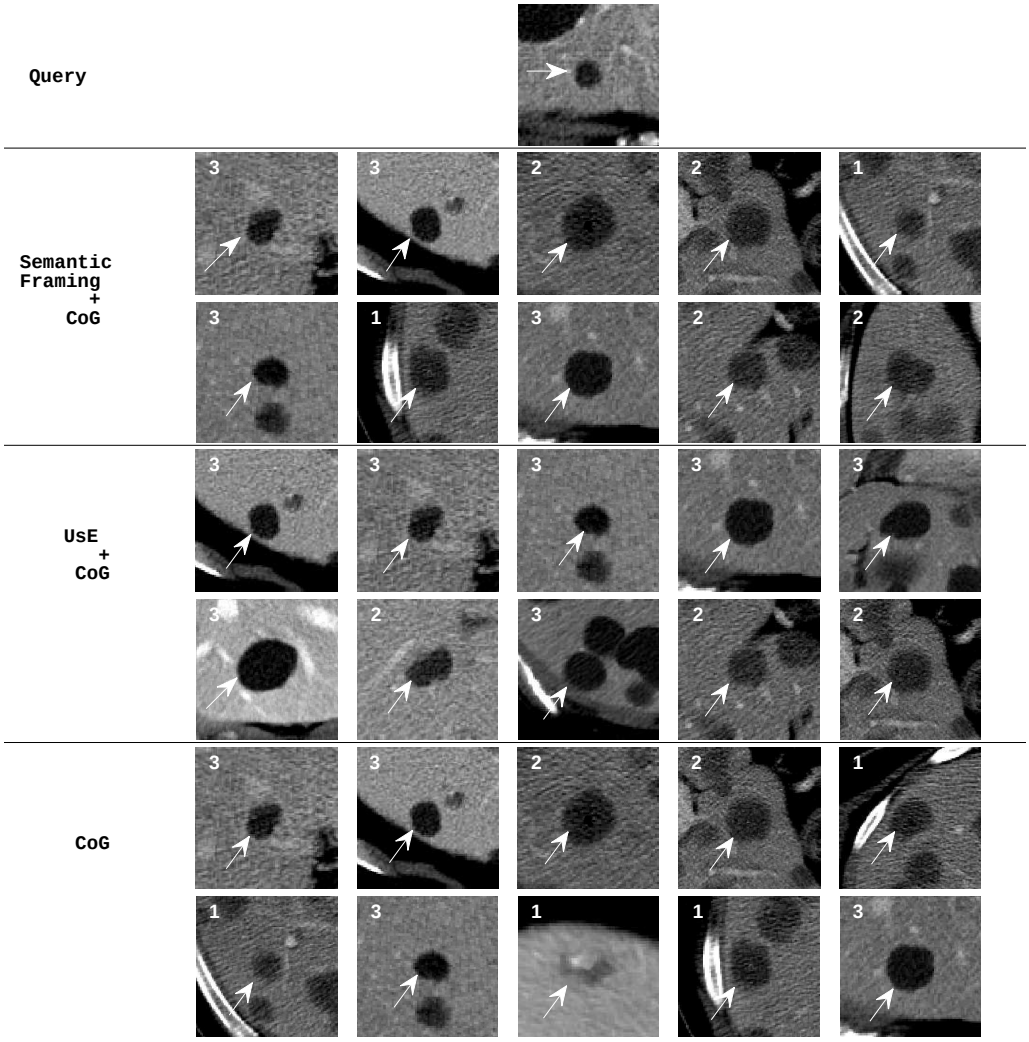

Figure 3: The retrieved lesions for a single query, in each of the three experiments, ordered row-wise according to the ranking results.The numbers on each retrieved lesion image is the ground truth grades (3: Similar, 2: Somewhat similar, 1: Dissimilar)

The objective of CaReRa is to perform *case* retrieval which requires the utilization of non-image metadata as well. Future work includes improving the semantic framing methodology by incorporating the ONLIRA ontology in full (currently, only its terms are used without any inter-relations), investigating context learning similarity ranking approaches to address the needs in typical scenarios where the users' intentions are unknown, expanding the metadata by including the non-image domain information to perform *case* retrieval within the CBCR concept.

7. ACKNOWLEDGEMENTS

We would like to thank Christopher F. Beaulieu, Daniel Rubin and Sandy Napel from Stanford University for providing the dataset. This work is in part supported by CaReRa Project (TÜBİTAK Project No: 110E264) and Bogazici University B.A.P (Project No: 5324).

8. REFERENCES

[1] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar. Content-based image retrieval in radiology: current status and future directions. *J. Digit. Imaging*, 24(2):208–222, 2011.

[2] T. M. Deserno, P. Welter, and A. Horsch. Towards a repository for standardized medical image and signal case data annotated with ground truth. *J. digit. imaging*, 25(2):213–226, 2012.

[3] B. Fischer, P. Welter, R. W. Günther, and T. M. Deserno. Web-based bone age assessment by content-based image retrieval for case-based reasoning. *Int. J. comp. assist. radio. and surgery*, 7(3):389–399, 2012.

[4] M. Gletsos, S. G. Mougiakakou, G. K. Matsopoulos, K. S. Nikita, A. S. Nikita, and D. Kelekis. A computer-aided diagnostic system to characterize ct

focal liver lesions: design and optimization of a neural network classifier. *IEEE Trans. Inf. Techno. Biomed.*, 7(3):153–162, 2003.

[5] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Trans. Inf. Systems (TOIS)*, 20(4):422–446, 2002.

[6] C. E. Kahn. Artificial intelligence in radiology: decision support systems. *Radiographics*, 14(4):849–861, 1994.

[7] N. Kökciyan, R. Türkay, S. Üsküdarlı, P. Yolum, B. Acar, and B. Bakır. Semantic description of liver ct images: An ontological approach. *submitted to IEEE J. Biomed. and Health Inf.*

[8] T. Majtner, D. Svoboda, et al. Extension of tamura texture features for 3d fluorescence microscopy. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 301–307. IEEE, 2012.

[9] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler, et al. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Int. J. med. inf.*, 73(1):1–24, 2004.

[10] 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 results1. *Radiology*, 256(1):243–252, 2010.

[11] D. Rubin. Creating and curating a terminology for radiology: ontology modeling and analysis. *J. Digit. Imaging*, 21(4):355–362, 2008.

[12] F. A. Sadjadi and E. L. Hall. Three-dimensional moment invariants. *IEEE Trans. Pattern Anal. Mach. Intell.*, (2):127–136, 1980.

[13] L. Shapiro, I. Atmosukarto, H. Cho, H. Lin, S. Ruiz-Correa, and J. Yuen. Similarity-based retrieval for biomedical applications. In *Case-Based Reasoning on Images and Signals*, pages 355–387. Springer, 2008.

[14] L. Shen and L. Bai. 3d gabor wavelets for evaluating spm normalization algorithm. *Med. image anal.*, 12(3):375, 2008.

[15] J. F. Sowa. Conceptual structures: information processing in mind and machine. 1983.

[16] V. Strela, P. N. Heller, G. Strang, P. Topiwala, and C. Heil. The application of multiwavelet filterbanks to image processing. *IEEE Trans. Image Process.*, 8(4):548–563, 1999.

[17] D. T. Features for Image Retrieval. Master's thesis, Human Language Technology and Pattern Recognition Group, RWTH Aachen University, Aachen, Germany, 2003.

[18] R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered Sets*, pages 445–470. Dordrecht-Boston: Reidel, 1982.

[19] J. Xu, S. Napel, H. Greenspan, C. F. Beaulieu, N. Agrawal, and D. Rubin. Quantifying the margin sharpness of lesions on radiological images for content-based image retrieval. *Med. Phys.*, 39:5405, 2012.