ARTICLE IN PRESS

Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations

Camille Kurtz [a,b,*], Christopher F. Beaulieu [a], Sandy Napel [a], Daniel L. Rubin [a,*]

[a] Department of Radiology, School of Medicine, Stanford University, USA
[b] LIPADE (EA 2517), University Paris Descartes, France

ARTICLE INFO

Article history:
Received 23 September 2013
Accepted 27 February 2014
Available online xxxx

Keywords:
Image retrieval
Semantic image annotation
Semantic-based distances
Ontologies
Computed tomographic (CT) images
Liver lesions

ABSTRACT

Computer-assisted image retrieval applications could assist radiologist interpretations by identifying similar images in large archives as a means to providing decision support. However, the semantic gap between low-level image features and their high level semantics may impair the system performances. Indeed, it can be challenging to comprehensively characterize the images using low-level imaging features to fully capture the visual appearance of diseases on images, and recently the use of semantic terms has been advocated to provide semantic descriptions of the visual contents of images. However, most of the existing image retrieval strategies do not consider the intrinsic properties of these terms during the comparison of the images beyond treating them as simple binary (presence/absence) features. We propose a new framework that includes semantic features in images and that enables retrieval of similar images in large databases based on their semantic relations. It is based on two main steps: (1) annotation of the images with semantic terms extracted from an ontology, and (2) evaluation of the similarity of image pairs by computing the similarity between the terms using the Hierarchical Semantic-Based Distance (HSBD) coupled to an ontological measure. The combination of these two steps provides a means of capturing the semantic correlations among the terms used to characterize the images that can be considered as a potential solution to deal with the semantic gap problem. We validate this approach in the context of the retrieval and the classification of 2D regions of interest (ROIs) extracted from computed tomographic (CT) images of the liver. Under this framework, retrieval accuracy of more than 0.96 was obtained on a 30-images dataset using the Normalized Discounted Cumulative Gain (NDCG) index that is a standard technique used to measure the effectiveness of information retrieval algorithms when a separate reference standard is available. Classification results of more than 95% were obtained on a 77-images dataset. For comparison purpose, the use of the Earth Mover's Distance (EMD), which is an alternative distance metric that considers all the existing relations among the terms, led to results retrieval accuracy of 0.95 and classification results of 93% with a higher computational cost. The results provided by the presented framework are competitive with the state-of-the-art and emphasize the usefulness of the proposed methodology for radiology image retrieval and classification.

1. Introduction

1.1. Context

Diagnostic radiologists need to maintain high interpretation accuracy while maximizing efficiency in the face of increasing volumes of images per study. They are now confronted with the challenge of efficiently and accurately interpreting cross-sectional studies that often contain thousands of images [1]. Currently, this is largely an unassisted and time-consuming process, and a radiologist's accuracy is established through training and experience. Despite this training, there is substantial variation in interpretation among radiologists [2], and accuracy varies widely [3]. A promising approach to maintain interpretative accuracy in this "deluge" of data is to integrate computer-based assistance into the image interpretation process. Many general-purpose image retrieval systems have been proposed in the literature [4]. Among these systems, an emerging technique that may assist radiology interpretation is content-based image retrieval (CBIR) [5]. This framework assists users in finding similar images within large collections of images. For medical purposes, the role of CBIR is powerful: in

* Corresponding authors. Address: LIPADE (EA 2517), Université Paris Descartes, 45 Rue des Saints-Pères, 75006 Paris, France.

E-mail addresses: camille.kurtz@parisdescartes.fr (C. Kurtz), dlrubin@stanford.edu (D.L. Rubin).

http://dx.doi.org/10.1016/j.jbi.2014.02.018
1532-0464/

addition to enable similarity-based indexing, it could provide computer-aided diagnostic support based on image content and on other meta-data associated with images.

The main idea of CBIR is to search for similar images based directly on their visual contents. Image retrieval is usually performed by image 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 [6]. In general, images are indexed using quantitative features extracted from regions of interest (ROI) of the images (*e.g.*, lesions) and focus on their contents (*e.g.*, shape, texture) [7]. Although these features are powerful to describe the image content in an automated fashion, they are often not discriminating enough to comprehensively characterize medical images. In addition, the performance of most CBIR systems is constrained by the low-level properties of these features because they cannot efficiently model the user's high-level expectations [8] (referred to as the semantic gap problem). Since this problem remains unsolved, research in image retrieval focuses on new methods to characterize the image content with higher level semantics, closer to that familiar to the user and potentially more useful in retrieving similar-appearing images [9].

In recent work on image retrieval that incorporates semantics, the images were characterized using a set of semantic terms [10–12] in a process referred to as "semantic annotation". Such terms can be directly derived from the terminology provided by the radiologists in their reports [13] or automatically predicted from computational imaging features [14]. The semantic terms can be used to describe a variety of information about the image content (*e.g.*, lesion shape, patterns of enhancement), and they are directly linked to the user's high-level understanding and descriptions of image features [15] (Fig. 1). These terms can improve diagnostic decision making by enabling radiologists to search databases of images for cases that are similar in terms of shared high-level imaging features to the cases which they are working on. Based on these considerations, incorporating semantic features into CBIR systems can be a promising attempt to bridge the semantic gap between the visual description of an image and its meaning [16]. However, most of the existing CBIR strategies do not consider the intrinsic semantic properties of the terms during the comparison of the images. Consequently, there is an unmet need that we propose to address by presenting a new framework that includes semantic features in images and that enables retrieval of similar images in large databases.

1.2. Motivations

When images are described using semantic terms, they are usually modeled as a set of terms, referred to as "bag-of-words" (BOW) [17]. BOW models have been successfully used in natural language processing [18] to capture a summary of the semantics of text based on word content. BOW models are represented as vectors

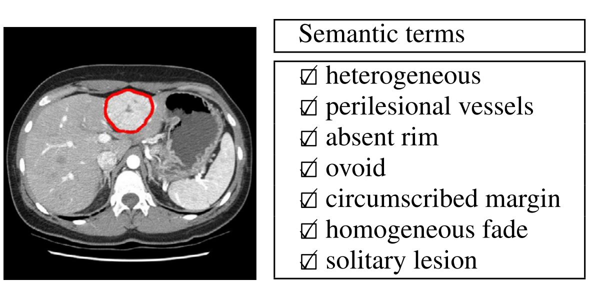

Fig. 1. A CT image of the liver annotated with semantic terms. The boundaries of the lesion are depicted in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of numerical (or binary) values, where each element represents the probability of occurrence (or presence/absence) of a term. Most of the classical BOW approaches assume that every term describing an image is independent of other features – that there are no intrinsic relations between the words that are contained in a bag. However, such an assumption is often oversimplified; while in text the independence assumption is reasonable, in the case of images annotated with diverse semantic features, this assumption is problematic [19]. The relations among terms are crucial since these features usually have a strong semantic correlation with each other (*e.g.*, relations between anatomy, imaging observations).

Consider an example of medical image clustering with three images $\mathcal{I}_{A-C}$. Each image contains a lesion in the liver. We use a basic vocabulary to describe the lesion shape appearing in these images: {*ovoid*, *round*, *irregular*}. The three images are annotated as presented: $\mathcal{I}_A$ is annotated with *ovoid*; $\mathcal{I}_B$ is annotated with *round*; and $\mathcal{I}_C$ is annotated with *irregular*. By considering a classical BOW strategy, these three images could be represented in a 3-dimensional space as $\mathcal{I}_A = (1,0,0)$, $\mathcal{I}_B = (0,1,0)$ and $\mathcal{I}_C = (0,0,1)$. Fig. 2(a) shows this representation in a classical Euclidean space. By considering this representation, it is difficult to group the three images into relevant clusters of interest because each image is equidistant to every other image. However, we know intuitively that the images $\mathcal{I}_A$ and $\mathcal{I}_B$ are more similar than the images $\mathcal{I}_A$ and $\mathcal{I}_C$ (resp. $\mathcal{I}_B$ and $\mathcal{I}_C$) since the terms *ovoid* and *round* are semantically closer than the terms *ovoid* and *irregular* (resp. *round* and *irregular*). Therefore, if we no longer hold the terms in the bag as orthogonal, it seems natural to "bend" the axis and thus enable easy clustering them into two groups (Fig. 2(b)).

This basic example highlights the need of considering the semantic relations of the terms for image retrieval purpose. Although several efforts have been conducted in computer vision [20,21] and medical imaging [22] to integrate semantics into image retrieval applications, most of the proposed approaches are dedicated to automatic annotation of the images with semantic terms, and they do not focus on the potential relations among the semantic terms during the retrieval step. Consequently, there is an opportunity to improve image retrieval applications by considering these semantic aspects. To this end, our approach to improving CBIR considers the semantic relations between the terms when assessing the distance between images described with BOW. Indeed, assessing the distance between vectors is the basis of determining the similarity in most medical information retrieval applications [23,24]. However, to enable the integration of such semantic relations into a distance function, we have to find solutions to three major problems: (1) how to model the relationship between the terms, (2) how to quantify a semantic proximity between terms, and (3) how to use these relations in computing the distance between images described as BOW.

Recent works in information retrieval [25] have shown that considering controlled vocabularies, such as ontologies, for image annotation can open up new research directions to deal with these problems. Ontologies can be used to model the relations between terms and can provide a solution to deal with our first issue. In addition, the problem of quantifying a semantic proximity between terms belonging to an ontology has been studied in the field of natural language processing [26]. Numerous kinds of measures have been proposed to assess the semantic similarity between terms and can be used as robust solutions to our second issue. Finally, in the domain of histogram comparison, new distances resting on hierarchical merging strategies have been proposed to consider the relative proximity between the bins avoiding bins correlation issues. The cornerstone of these distances is to consider the (intrinsic) multilevel semantic correlations between the distributions modeled by the histograms. Such distances could then be adapted to consider semantic similarity between terms when

Please cite this article in press as: Kurtz C et al. A hierarchical knowledge-based approach for retrieving similar medical images described with semantic annotations. J Biomed Inform (2014), http://dx.doi.org/10.1016/j.jbi.2014.02.018