Computerized Medical Imaging and Graphics 49 (2016) 37–45

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Adapting content-based image retrieval techniques for the semantic annotation of medical images

Ashnil Kumar [a,c,*], Shane Dyer [b], Jinman Kim [a,c], Changyang Li [a,c], Philip H.W. Leong [b,c], Michael Fulham [a,c,d,e], Dagan Feng [a,c,f]

[a] School of Information Technologies, University of Sydney, Australia
[b] School of Electrical and Information Engineering, University of Sydney, Australia
[c] Institute of Biomedical Engineering and Technology, University of Sydney, Australia
[d] Department of Molecular Imaging, Royal Prince Alfred Hospital, Sydney, Australia
[e] Sydney Medical School, University of Sydney, Australia
[f] Med-X Research Institute, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:
Received 10 June 2015
Received in revised form
10 December 2015
Accepted 14 January 2016

Keywords:
Image annotation
Content-based image retrieval
Computed tomography
Liver
ImageCLEF

ABSTRACT

The automatic annotation of medical images is a prerequisite for building comprehensive semantic archives that can be used to enhance evidence-based diagnosis, physician education, and biomedical research. Annotation also has important applications in the automatic generation of structured radiology reports. Much of the prior research work has focused on annotating images with properties such as the modality of the image, or the biological system or body region being imaged. However, many challenges remain for the annotation of high-level semantic content in medical images (e.g., presence of calcification, vessel obstruction, etc.) due to the difficulty in discovering relationships and associations between low-level image features and high-level semantic concepts. This difficulty is further compounded by the lack of labelled training data. In this paper, we present a method for the automatic semantic annotation of medical images that leverages techniques from content-based image retrieval (CBIR). CBIR is a well-established image search technology that uses quantifiable low-level image features to represent the high-level semantic content depicted in those images. Our method extends CBIR techniques to identify or retrieve a collection of labelled images that have similar low-level features and then uses this collection to determine the best high-level semantic annotations. We demonstrate our annotation method using retrieval via weighted nearest-neighbour retrieval and multi-class classification to show that our approach is viable regardless of the underlying retrieval strategy. We experimentally compared our method with several well-established baseline techniques (classification and regression) and showed that our method achieved the highest accuracy in the annotation of liver computed tomography (CT) images.

1. Introduction

1.1. Motivation and aims

Medical imaging is a fundamental component of modern healthcare with roles in patient diagnosis, treatment planning, and assessment of response to therapy. A direct consequence of this is the rise in medical imaging informatics research, including content-based image retrieval [1,2], modality-classification and case-based retrieval [3], classification [4,5], and annotation [5–7]. Semantic image annotation is also emerging as a research question, in which the main research challenge is to detect subtle differences in low-level image features and to relate them to higher-level labels derived from a standard terminology. Ultimately the goal is to apply the annotation technologies for the automatic generation of structured imaging reports [8,9].

Annotation is also considered to be a prerequisite for semantic medical search engines that enable radiologists to find medical images, reports, and associated publications more efficiently [7]. Automatic semantic annotation is needed because it is difficult,

* Corresponding author at: School of Information Technologies, University of Sydney, Building J12, NSW 2006, Australia. Tel.: +61 2 9036 9805; fax: +61 2 9351 3838.
E-mail addresses: ashnil.kumar@sydney.edu.au (A. Kumar), sdye9175@uni.sydney.edu.au (S. Dyer), jinman.kim@sydney.edu.au (J. Kim), changyang.li@sydney.edu.au (C. Li), philip.leong@sydney.edu.au (P.H.W. Leong), michael.fulham@sydney.edu.au (M. Fulham), dagan.feng@sydney.edu.au (D. Feng).

http://dx.doi.org/10.1016/j.compmedimag.2016.01.001
0895-6111/

time-consuming and expensive to manually annotate the rich contents of these items. The annotation and image markup use case of the caBIG project [10], which described a software library that could be used for the annotation of large collections of images, provides an example of the ponderous nature of manual annotation processes. Wennerberg et al. [7] improved the efficiency of this manual annotation process using an ontology modularisation tool that identifies and ranks fragments of an ontology that are relevant to the annotation task; this relevance is based upon the specific domain (e.g., lymphoma) and hierarchical relationships of terms already annotated. However, these manual annotation approaches require physicians to subjectively determine the labels that are relevant to a particular image based on the physicians' expertise and prior experience.

In contrast, automatic image annotation is conducted on the basis of quantifiable image features. The combination of features present in each image suggests the annotations that are relevant. Many existing approaches described in the summary paper by Deselaers et al. [11] only annotated the images with the properties of the image, such as the image modality, body orientation, body region and biological system being examined. Setia et al. [5] extracted local feature descriptors from the most salient (interesting) points on each image to capture the geometric relationships present in the image; a hierarchical classification method was used to annotate each image by the image properties listed earlier. In a similar application, Tommasi et al. [6] proposed a method that extracted global and local features using three classification strategies that emphasised feature fusion at different stages of the annotation process. Ko et al. [12] presented a method that utilised a random forest classifier together with a predefined body relation graph to identify and annotate the body region shown in the image.

A more difficult objective is to annotate the images with clinically relevant *content*, such as the presence of calcification, mass effect, etc. In the general (i.e., non-medical) domain, image annotation tasks have moved rapidly from object identification to sentence generation, where the aim is to describe the images through words, in the same way in which a human witness might describe a scene that they have observed; several such methods have been described in a recent summary paper [13]. Kulkarni et al. [14] used computer vision based object detection to construct a graph of the objects and labeled the graph based upon statistics mined from large corpora of descriptive text; the labels and graph relationships could then be used to generate descriptive sentences.

One of the major hurdles in achieving this objective for medical images is that there are likely to be thousands of semantic labels to learn and often very few labeled training samples [15]. Thus a major challenge of such research is the development of categorisation and annotation techniques that are less hindered by lack of training samples [16]. To reduce problems caused by lack of training data, Gimenez et al. [17] avoided classification methods and instead annotated liver CT images using logistic regression, through the least absolute shrinkage and selection operator (LASSO). However, their method only annotated binary semantic outcomes that could be presented by positive or negative observations, e.g., whether or not a lesion was homogeneous. In a follow-up study, Depeursinge et al. [18] learned semantic terms describing the visual appearance of liver lesions derived from a linear combination of multi-scale wavelet features. This allowed their method to model each annotation at the most relevant image scale. The method predicted the probability that a particular semantic description (e.g., irregular lesion margin) was applicable to the lesion in the image but did not annotate the effects on anatomical structures, e.g., the proximity of the lesion to the hepatic vasculature.

The recognition of image content also falls within the scope of another important area of medical imaging informatics research called content-based image retrieval (CBIR) [1]. In CBIR, low-level visual features such as intensity, texture, shape, and the spatial arrangement of objects are used to determine which images are similar to a given query [19]. A key challenge for CBIR is the *semantic gap*, which is the difference between machine-computed similarity and a human's interpretation of similarity [19]. Many different CBIR algorithms have been investigated for this purpose; a summary can be found in the recent review by Kumar et al. [2]. Well-established CBIR techniques are therefore designed to relate low-level image features to higher-level semantic concepts. We hypothesise that the problem of automatic semantic image annotation could be addressed in a related fashion, by adapting the ability of CBIR techniques to leverage low-level image features in the search for images with similar high-level semantic concepts.

Thus in this paper, we present a method for the automatic annotation of medical images that is derived from CBIR techniques. Given an image to annotate, we propose to identify or retrieve a collection of semantically similar images that have already been labelled and use this collection to determine the best semantic annotations for the unlabelled image. Our annotation method is designed for limited training data compared to the number of annotations that need to be automatically recognised. We suggest that the technique would be applicable regardless of the underlying retrieval strategy and therefore we describe two ways of identifying the best annotations: either through multi-class classification and nearest-neighbour search, both of which are well-established CBIR methods. We evaluated our work on the annotation of liver CT images by comparing our annotation method to several other well-established techniques. We also compared our method to the state-of-the-art techniques submitted to the Imaging track of the Conference and Labs of the Evaluation Forum (ImageCLEF) [20] Liver Annotation Challenge [21]; the outcomes were reported at the CLEF workshop [22]. In this paper we expand upon the report by including: (i) detailed definitions of the classification and nearest neighbour methods for annotation, and (ii) a more comprehensive evaluation, which includes comparison with well-established techniques that were not submitted to the ImageCLEF Liver Annotation Challenge.

1.2. Terminology and notation

We employ the following terminology in the remainder of this paper. A *question* refers to a specific annotation task, i.e., an element of the structured report that needs to be automatically filled. A *label* is an annotation that could possibly be assigned to a question. An *answer* is the label that our method automatically assigns to the question based on the analysis of the image features; the answer is chosen from a set of labels that are unique to each question. The term *query* refers to a single un-annotated image volume that will be annotated using our approach.

We also use the following notation. Let Ω be a question and $\mathcal{L}_\Omega$ be the set of labels for Ω with $\left|\mathcal{L}_\Omega\right| = l$. During annotation, we also let $\mathcal{L}_\Omega^+ \subseteq \mathcal{L}_\Omega$ denote a possible set of answers (needed only in case of ties) and $L \in \mathcal{L}_\Omega^+$ denotes the final answer. Note that since only one label is chosen as the final answer (i.e., $\left|L\right| = 1$) then $L = \mathcal{L}_\Omega^+ \Leftrightarrow \left|\mathcal{L}_\Omega^+\right| = 1$ (there were no ties).

2. Materials and methods

2.1. Dataset

We used a public dataset of volumetric (3D) computed tomography (CT) images of the liver from the ImageCLEF 2014 Liver Annotation Challenge [21]. The dataset contained 50 CT volumes cropped to the region around the liver; the volumes had varied resolutions (x: 190–308 pixels, y: 213–387 pixels, slices: 41–588)