Computerized Medical Imaging and Graphics 39 (2015) 14–26

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

Improved medical image modality classification using a combination of visual and textual features

Ivica Dimitrovski [a], Dragi Kocev [b], Ivan Kitanovski [a], Suzana Loskovska [a], Sašo Džeroski [b]

[a] Faculty of Computer Science and Engineering, University Ss. Cyril and Methodius, Skopje, Macedonia
[b] Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 2 June 2014
Accepted 5 June 2014

Keywords:
Image modality classification
Visual image descriptors
Feature fusion

ABSTRACT

In this paper, we present the approach that we applied to the medical modality classification tasks at the ImageCLEF evaluation forum. More specifically, we used the modality classification databases from the ImageCLEF competitions in 2011, 2012 and 2013, described by four visual and one textual types of features, and combinations thereof. We used local binary patterns, color and edge directivity descriptors, fuzzy color and texture histogram and scale-invariant feature transform (and its variant opponentSIFT) as visual features and the standard bag-of-words textual representation coupled with TF-IDF weighting. The results from the extensive experimental evaluation identify the SIFT and opponentSIFT features as the best performing features for modality classification. Next, the low-level fusion of the visual features improves the predictive performance of the classifiers. This is because the different features are able to capture different aspects of an image, their combination offering a more complete representation of the visual content in an image. Moreover, adding textual features further increases the predictive performance. Finally, the results obtained with our approach are the best results reported on these databases so far.

1. Introduction

Large collections of medical images have become a valuable source of knowledge, taking an important role in education, medical research and clinical decision making. One of the main problems is that the size of the collections is in constant growth due to the increasing availability of imaging equipment in hospitals. Average-sized radiology departments now produce several tera-bytes of data annually. This generates huge repositories of valuable information, which in many cases is difficult to process and manage appropriately. This prompts for development of tools for efficient and effective access to this type of information.

Medical image databases are typically accessed via textual information through the standard Picture Archiving and Communication System (PACS) [1,2]. PACS integrates imaging modalities and interfaces with hospital and departmental information systems to manage storage and distribution of images to medical personnel, researchers, clinics, and imaging centers. The task of indexing and cataloging these collections has been traditionally performed manually. This is an expensive and time-consuming procedure, and it is also prone to errors. Consequently, there is a strong need for automated indexing of medical image collections in order to improve the ability to search for and retrieve relevant images [3].

Medical image retrieval systems have traditionally been text-based, relying on the annotation or captions associated with the images as the input to the retrieval system. In the last few decades, several advancements in the area of content-based image retrieval (CBIR) have been made [4,5]. CBIR systems have had some success in fairly constrained medical domains, such as pathology [6], head MRIs [7], lung CTs [8], and mammograms [9]. Furthermore, combining both textual and visual techniques improves the retrieval performance over using them individually [3,10]. The queries, in that case, consist of a textual part (i.e., textual sub-query) and/or sample images (i.e., visual sub-query). For example, the queries could contain information about patients' demographics, a limited set of symptoms and medical examination results including imaging studies.

Medical image databases used for retrieval or for teaching purposes often contain images of many different modalities, such as X-ray, CT scan, ultrasound, etc. An additional complication is that these images are typically taken under different conditions and the accuracy of their annotations is variable and inconsistent [11]. This is especially true for images found in various on-line resources, including those that access the on-line content of journals.

E-mail addresses: Ivica.Dimitrovski@finki.ukim.mk (I. Dimitrovski), Dragi.Kocev@ijs.si (D. Kocev), Ivan.Kitanovski@finki.ukim.mk (I. Kitanovski), Suzana.Loskovska@finki.ukim.mk (S. Loskovska), Saso.Dzeroski@ijs.si (S. Džeroski).

http://dx.doi.org/10.1016/j.compmedimag.2014.06.005
0895-6111/

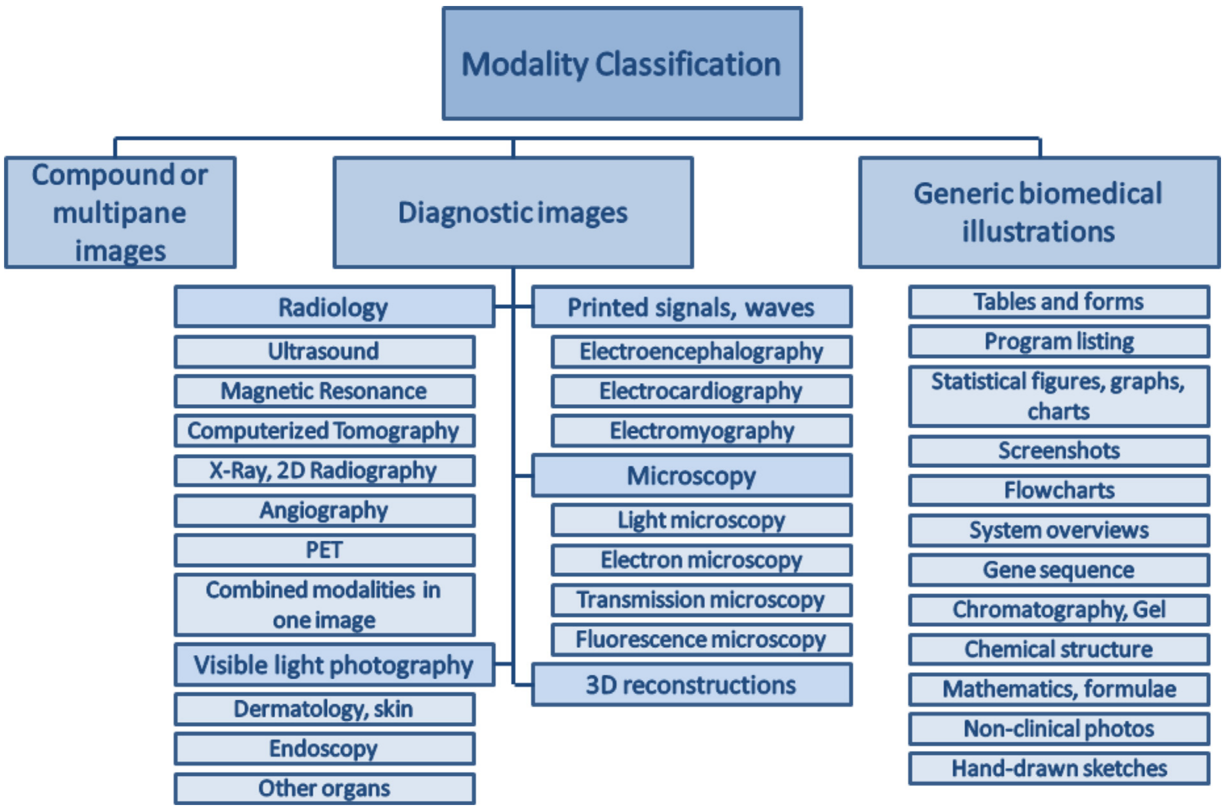

Fig. 1. The classification hierarchy used for the ImageCLEF 2012 and 2013 competitions. The image is taken from http://www.imageclef.org/2013/medical.

Image modality is a fundamental visual characteristic of an image and can be exploited for improving retrieval performance. However, the annotations or captions associated with images often do not capture information about the modality. The DICOM header contains tags to decode the body part examined, the patient position and the modality [12]. Some of the tags are automatically set by the digital imaging system according to the imaging protocol used to capture the pixel data. Other tags are set manually by the physicians or radiologists during routine documentation. This procedure cannot always be considered very reliable, since it frequently happens that some entries are either missing, false, or do not describe the anatomic region precisely [13].

The medical retrieval task in ImageCLEF has provided both a forum as well as test image collections to benchmark image retrieval techniques. Over the years, our group has participated in different subtasks of this medical task including the subtasks of automatic medical image annotation [14], medical modality classification, compound figure separation, ad-hoc image-based retrieval and case-based retrieval [3,10]. In this paper, we present in detail the results of applying our approach to modality classification for the competitions organized in 2011, 2012 and 2013.

In the modality classification task at the ImageCLEF competition, the examples are images from medical articles. The goal is to correctly classify the modality of the images using the visual information from the images and the text from the article where this image is encountered. In this work, we extensively compare the performance of 4 different techniques for feature extraction from images: local binary patterns (LBP) [15], the color and edge directivity descriptor (CEDD) [16], fuzzy color and texture histograms (FCTH) [17] and the scale-invariant feature transform (SIFT) with its variant opponentSIFT (OSIFT) [18,19]. Next, we evaluate the performance of textual features extracted from the text surrounding the images by using the standard bag-of-words representation together with the TF-IDF weighting [20].

The main focus of this work is to first explore which visual features extraction technique captures the most relevant information about the medical image modality. Second, we investigate whether the combination of the different visual features improves the predictive performance. Next, we compare the performance of visual and textual features in the context of medical image modality classification. Finally, we investigate whether combining visual and textual features improves the performance and yields state-of-the-art performance.

The reminder of this paper is organized as follows. Section 2 presents the task of modality classification at ImageCLEF competitions. The visual and textual feature extraction techniques are described in Section 3. The specific experimental setup used to evaluate the feature extraction techniques is outlined in Section 4. Section 5 discusses the results from the experiments and compares the different evaluation scenarios. Finally, Section 6 states our conclusions.

2. The task of modality classification

For medical retrieval purposes, imaging modality is an important aspect of the image. In user studies, clinicians have indicated that modality is one of the most important filters that they would like to be able to limit their search by [3]. The usage of modality information often significantly improves the retrieval results.

The ImageCLEF medical modality classification task is a standardized benchmark for systems that automatically classify medical image modality from PubMed journal articles. This task was first introduced in ImageCLEF 2010, when the total number of modalities was eight. In ImageCLEF 2011, the number of modalities was expanded to 18. The ImageCLEF 2012 and 2013 dataset have 31 classes (same number of classes and same classification hierarchy): However, in ImageCLEF 2013, a larger number of compound/multipane images (i.e., images that contain figures of several