Journal of Biomedical Informatics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Histology image search using multimodal fusion

Juan C. Caicedo [a,*,1], Jorge A. Vanegas [b], Fabian Paez [b], Fabio A. González [b]

[a] University of Illinois at Urbana–Champaign, IL, USA
[b] MindLab Research Laboratory, Universidad Nacional de Colombia, Bogotá, Colombia

ARTICLE INFO

Article history:
Received 18 June 2013
Accepted 30 April 2014
Available online xxxx

Keywords:
Histology
Digital pathology
Image search
Multimodal fusion
Visual representation
Semantic spaces

ABSTRACT

This work proposes a histology image indexing strategy based on multimodal representations obtained from the combination of visual features and associated semantic annotations. Both data modalities are complementary information sources for an image retrieval system, since visual features lack explicit semantic information and semantic terms do not usually describe the visual appearance of images. The paper proposes a novel strategy to build a fused image representation using matrix factorization algorithms and data reconstruction principles to generate a set of multimodal features. The methodology can seamlessly recover the multimodal representation of images without semantic annotations, allowing us to index new images using visual features only, and also accepting single example images as queries. Experimental evaluations on three different histology image data sets show that our strategy is a simple, yet effective approach to building multimodal representations for histology image search, and outperforms the response of the popular late fusion approach to combine information.

1. Introduction

Digital pathology makes it easy to exchange histology images and enables pathologists to rapidly study multiple samples from different cases without having to unpack the glass [1]. The increasing adoption of digital repositories for microscopy images results in large databases with thousands of records, which may be useful to supporting the decision making process in clinical and research activities. However, in modern hospitals and health care centers, the number of images to keep track of is beyond the ability of any specialist. A very promising direction to realize the potential of these collections is through efficient and effective tools for *image search*. For instance, when a new slide is being observed, a camera coupled to the microscope can capture the current view, send the picture to the retrieval system, and show results on a connected computer. These results can help to clarify structures in the observed image, explore previous cases and, in general, may allow clinicians and researchers to explore large collections of records previously evaluated and diagnosed by other physicians.

The *query-by-example* paradigm for image search—when the user's query is an example image with no annotations—has a number of potential applications in medicine and clinical activities [2]. The main challenge when implementing such a system consists of correctly defining the matching criteria between query images and database images. The standard approach for content-based retrieval in image collections relies on using similarity measures between low-level visual features to perform a nearest-neighbor search [3]. The problem of this approach is that these characteristics usually fail to capture high-level semantics of images, a problem known as the semantic gap [4]. Different methods to bridge this gap have been proposed to build a model that connects low-level features with high-level semantic content, such as automatic image annotation [5] and query by semantic example [6]. These methods represent images in a semantic space spanned by keywords, so a nearest neighbors search in that space retrieves semantically related images. Approaches like these have also been investigated for histology image search [7–9].

Image search systems based on a semantic representation have been shown to outperform purely visual search systems in terms of Mean Average Precision (MAP) [3]. However, these approaches may lose the notion of visual similarity among images since the search process ends up relying entirely on high level descriptions of images. The ranking of search results is based on potentially relevant keywords, ignoring useful appearance clues that are not described by index terms. In a clinical setting, visual information plays an important role for searching histology images, which

* Corresponding author. Current address: Computer Science Department, University of Illinois at Urbana-Champaign, 201 N Goodwin Ave, Urbana, IL 61801, USA.
E-mail addresses: caicedo@illinois.edu (J.C. Caicedo), javanegasr@unal.edu.co (J.A. Vanegas), fmpaezri@unal.edu.co (F. Paez), fagonzalezo@unal.edu.co (F.A. González).
[1] Work done while at Universidad Nacional de Colombia.

http://dx.doi.org/10.1016/j.jbi.2014.04.016
1532-0464/

ultimately reveals the biological evidence for the decision making process in clinical activities. We consider that both visual content and semantic data are complementary sources of information that may be combined to produce high quality search results.

Multimodal fusion has emerged as a very useful approach to combine different signals with the purpose of making certain semantic decisions in automated systems. We refer the reader to [10] for a comprehensive survey of multimodal fusion in various multimedia applications. For image indexing in particular, multimodal fusion consists of combining visual and semantic data. Several methodologies have recently been proposed to model the relationships between these two data modalities, with the goal of constructing better image search systems. Two main strategies may be identified to achieve the combination of both data modalities: (1) *early fusion* [11], to build a combined representation of images before the ranking procedure and (2) *late fusion* [12], to combine similarity measures during the ranking procedure. One of the advantages of early fusion over late fusion is that the former often benefits from explicitly modeling the relationships between the two data modalities, instead of simply using them as separate opinions. However, this clearly requires a significant effort in understanding and extracting multimodal correspondences.

In this work, we propose a novel method for indexing histology images using an early multimodal fusion approach, that is, combining the two data modalities in a single representation to generate the ranking directly in such a space. The proposed methods use semantic annotations as an additional data source that represents images in a vector space model. Then, matrix-factorization-based algorithms are used to find the relationships between data modalities, by learning a function that projects visual data to the semantic space and the other way around. We take advantage of this property by fusing both data modalities in the same vector space, obtaining as a result the combined representation of images.

A systematic experimental evaluation was conducted on three different histology image databases. Our goal is to validate the potential of various image search techniques to understand the strengths and weaknesses of visual, semantic and multimodal indexing in histology image collections. We focus our evaluation on two performance measures commonly used for information retrieval research: Mean Average Precision (MAP); and Precision at the first 10 results of the ranked list (P@10), for early precision. We observed that semantic approaches are very good at maximizing MAP, while visual search is a strong baseline for P@10, revealing a trade-off in performance when using one or the other representation. This also confirms the importance of combining both data modalities.

Our approach combines multimodal data using a convex combination of the visual and semantic information, resulting in a continuous spectrum of multimodal representations and allowing us to explore various mixes from purely visual to purely semantic representations as needed. This is similar in spirit to late fusion, which allows the setting of weights to scores produced by each modality. However, our study shows significant improvement in performance when building an explicitly fused representation, instead of considering modalities as separate voters for the rank of images. We also found that multimodal fusion can balance a trade-off between maximizing MAP and early precision, demonstrating the potential to improve the response of histology image retrieval systems.

1.1. Overview

This work proposes an indexing technique for image search, using both visual image content and associated semantic terms. Fig. 1 illustrates a pipeline for image search in a clinical setting, which involves a physician or expert pathologist working with a microscopy equipment with digital image acquisition capabilities or in a virtual microscopy system. Through an interactive mechanism, the user can ask the system to take a picture of the current view, and send a query to the image search system. The system has a pre-computed fused representation of images in the database. A ranking algorithm is used to identify the most relevant results in the database, which are retrieved and presented to the user.

The main goal of the system is to support to clinicians during the decision making process by providing relevant associated information. The ability to find related cases among past records in a database has the potential to improve the quality of health care using an evidence-based reasoning approach. Historic archives in a hospital comprise a knowledge base reflecting its institutional experience and expertise, and it can be used to enhance the daily medical practice.

This paper focuses on two important aspects of the entire pipeline: (1) strategies for constructing the index based on a multimodal fused representation and (2) an empirical evaluation of different strategies for histology image search using collections of real diagnostic images. The main contribution of our work is a novel method for combining visual and semantic data in a fused image representation, using a computationally efficient strategy that outperforms the popular late-fusion approach, and that balances the trade-off between visual and semantic data. While the applicability of the proposed model may be extended to general image collections beyond histology images, the second contribution of this work is an extensive evaluation on histology images, since a straightforward application of image retrieval techniques may not result in an optimal outcome. Part of our experimental evaluation shows that off-the-shelf indexing methods such as latent semantic indexing and late fusion do not always exploit specific characteristics of histology images.

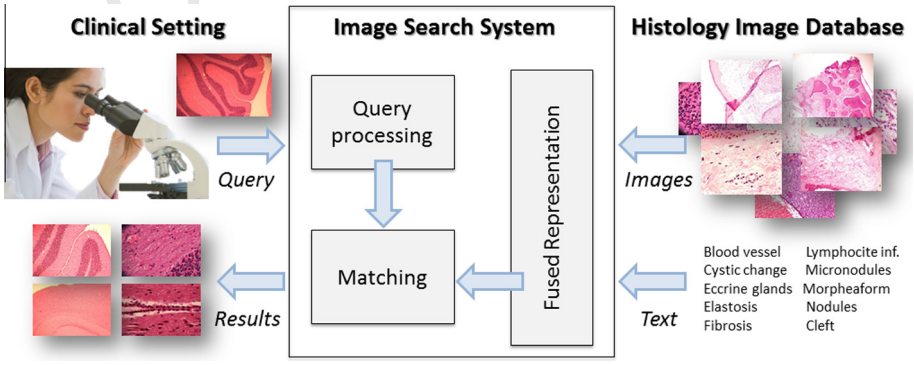

Fig. 1. Overview of the image search pipeline. Images acquired in a clinical setting are used as example queries. The system processes and matches queries with entries in a multimodal index, which represent images and text in the database. Results are returned to support the decision making process.

Q1 Please cite this article in press as: Caicedo JC et al. Histology image search using multimodal fusion. J Biomed Inform (2014), http://dx.doi.org/10.1016/j.jbi.2014.04.016