Medical Image Analysis 29 (2016) 47–64

Contents lists available at ScienceDirect

Medical Image Analysis

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

Automatic detection of referral patients due to retinal pathologies through data mining[☆]

Gwenolé Quellec[a,*], Mathieu Lamard[a,b], Ali Erginay[c], Agnès Chabouis[c], Pascale Massin[c], Béatrice Cochener[a,b,d], Guy Cazuguel[a,e]

[a] Inserm, UMR 1101, SFR ScInBioS, F-29200 Brest, France
[b] Univ Bretagne Occidentale, F-29200 Brest, France
[c] Service d'Ophtalmologie, Hôpital Lariboisière, APHP, F-75475 Paris, France
[d] Service d'Ophtalmologie, CHRU Brest, F-29200 Brest, France
[e] Institut Mines-Telecom; Telecom Bretagne; UEB; Dpt ITI, F-29200 Brest, France

ARTICLE INFO

Article history:
Received 6 July 2015
Revised 17 December 2015
Accepted 18 December 2015
Available online 29 December 2015

Keywords:
Retinal pathologies
Anomaly detection
Bag of Visual Words model
Data mining

ABSTRACT

With the increased prevalence of retinal pathologies, automating the detection of these pathologies is becoming more and more relevant. In the past few years, many algorithms have been developed for the automated detection of a specific pathology, typically diabetic retinopathy, using eye fundus photography. No matter how good these algorithms are, we believe many clinicians would not use automatic detection tools focusing on a single pathology and ignoring any other pathology present in the patient's retinas. To solve this issue, an algorithm for characterizing the appearance of abnormal retinas, as well as the appearance of the normal ones, is presented. This algorithm does not focus on individual images: it considers examination records consisting of multiple photographs of each retina, together with contextual information about the patient. Specifically, it relies on data mining in order to learn diagnosis rules from characterizations of fundus examination records. The main novelty is that the content of examination records (images and context) is characterized at multiple levels of spatial and lexical granularity: 1) spatial flexibility is ensured by an adaptive decomposition of composite retinal images into a cascade of regions, 2) lexical granularity is ensured by an adaptive decomposition of the feature space into a cascade of visual words. This multigranular representation allows for great flexibility in automatically characterizing normality and abnormality: it is possible to generate diagnosis rules whose precision and generalization ability can be traded off depending on data availability. A variation on usual data mining algorithms, originally designed to mine static data, is proposed so that contextual and visual data at adaptive granularity levels can be mined. This framework was evaluated in *e-ophtha*, a dataset of 25,702 examination records from the OPHDIAT screening network, as well as in the publicly-available Messidor dataset. It was successfully applied to the detection of patients that should be referred to an ophthalmologist and also to the specific detection of several pathologies.

1. Introduction

Retinal pathologies are responsible for millions of blindness cases worldwide. According to the World Health Organization, 4.5 million people are blind due to glaucoma, 3.5 million people are blind due to age-related macular degeneration and 2 million people are blind due to diabetic retinopathy.[1] Early diagnosis is the key to slowing down the progression of these diseases and therefore preventing the occurrence of blindness.

The leading modality for detecting retinal pathologies is color fundus photography. In recent years, many algorithms have been developed for the automatic analysis of color fundus photographs (Winder et al., 2009; Abràmoff et al., 2010a; Trucco et al., 2013). The goal usually is to detect the first appearing signs of a target pathology, such as microaneurysms in the case of diabetic retinopathy (Niemeijer et al., 2010). Based on these detections, an

☆ This paper was recommended for publication by Nicholas Ayache.

* Corresponding author at: LaTIM, Bâtiment 1, CHRU Morvan, 2, Av. Foch, 29609 Brest Cedex, France. Tel.: +33 2 98 01 81 29; fax: +33 2 98 01 81 24.

E-mail address: gwenole.quellec@inserm.fr, gwenole.quellec@live.com (G. Quellec).

[1] www.who.int/blindness/causes/priority.

http://dx.doi.org/10.1016/j.media.2015.12.006
1361-8415/

automated diagnosis can be made: is the target pathology present in the patient or not? Such automated diagnoses have proven to be as efficient as a second expert opinion, compared to the initial expert opinion (Abràmoff et al., 2010b, 2013). One envisioned application for such algorithms is the automatic detection of patients that should be referred to an ophthalmologist (Abràmoff et al., 2013; Decencière et al., 2013). However, an automatic diagnosis tool that can only diagnose one pathology is less likely to be used in that context: ophthalmologists would want to look at images themselves to make sure no other pathology is present. We believe the ability to detect as many anomalies as a human expert is required to allow widespread use of an automated diagnosis tool. In order to achieve this goal, one solution would be to define a specific detector for each possible anomaly. But due to the large range of retinal pathologies, this solution is too complex. A simpler solution is to define a general framework for characterizing the appearance of abnormal examination records or, alternatively, the appearance of the normal ones.

A few solutions have been presented in the past to characterize the appearance of abnormal examination records. These solutions typically rely on Multiple Instance Learning (MIL) (Amores, 2013). The purpose of MIL is to identify a local image pattern that can explain a diagnosis assigned to the case as a whole (Amores, 2013). Extensions to the identification of multiple local patterns have been presented (Foulds and Frank, 2010; Quellec et al., 2012). Recently, MIL has been applied to eye fundus examinations (Yu et al., 2004; Quellec et al., 2012; Venkatesan et al., 2012). Besides MIL, a few anomaly detectors have been presented for specific regions of the retina: the optic disk (Kavitha and Ramakrishnan, 2010; Zhu et al., 2014) and the retinal vasculature (Kavitha and Ramakrishnan, 2010). As an alternative, a few solutions have also been presented to characterize the appearance of normal cases. One promising solution is to build a statistical atlas describing the normal range of image intensities at every location of the retina (Lee et al., 2010). After registering an input image to this statistical atlas, anomalies can be detected by measuring the local deviation from the statistical atlas (Ali et al., 2013; Quellec et al., 2010b).

A data mining solution is presented in this paper. Data mining is used to characterize the appearance of abnormal examination records or, alternatively, the appearance of the normal ones, based on a large dataset of eye fundus examination records. Specifically, it allows the identification of “multimedia patterns” that are visible in examination records marked as abnormal, but not in examination records marked as normal. Or, alternatively, it allows the identification of multimedia patterns characterizing examination records marked as normal. In this paper, a multimedia pattern denotes either 1) a feature vector extracted locally from an image, 2) a conjunction of feature vectors extracted locally from one or several images, or 3) a conjunction of local feature vectors and contextual information about the patient. For improved performance, the localization of local feature vectors is taken into account. Therefore, it is possible to extract multimedia rules of this kind: “if the patient is a female younger than 40 and a pattern similar to this example was found within one millimeter of one of her optic disks and at least two patterns similar to that other example were found within 2 millimeters of one of her foveae, then she should be addressed to an ophthalmologist”. This solution is more general than any other existing solution and it makes the most of every available information. Three challenges need to be addressed:

1. relevant local feature vectors need to be extracted,
2. these feature vectors need to be localized with respect to retinal landmarks,
3. image-derived features and contextual information need to be fused.

Regarding the first challenge (*local feature vector extraction*), the Bag of Visual Words (BoVW) model (Sivic and Zisserman, 2003) is currently the most popular solution for characterizing an image based on its local appearance. This solution is fast and is scalable to large datasets (Ji et al., 2013; Liu et al., 2014). Therefore, it is well suited for a data mining task. Recently, the BoVW model has been applied to retinal images (Rocha et al., 2012; van Grinsven et al., 2013; Pires et al., 2013; 2014). Regarding the second challenge (*localization*), several algorithms have been presented for the automatic detection of the optic disk and the fovea in eye fundus photographs (Li and Chutatape, 2004; Niemeijer et al., 2009a; Gegundez-Arias et al., 2013). For improved spatial localization, algorithms have also been presented for the automatic formation of a composite image, or “mosaic”, per eye (Can et al., 2002; Choe et al., 2006; Li et al., 2008; Wei et al., 2009; Abràmoff et al., 2012). However, these algorithms assume that the photographer indicated which images correspond to the left eye and which images correspond to the right eye, which is generally not the case. Regarding the third challenge (*information fusion*), a few solutions have been presented in recent years for combining heterogeneous information extracted from several photographs in one examination record (Niemeijer et al., 2009b). We have developed alternative solutions, which are more general in the sense that they also take into account contextual information about the patient (structured demographic and clinical data) (Quellec et al., 2009, 2010a, 2011; Decencière et al., 2013). In these solutions, image-derived features and contextual information are combined through data mining (early fusion) or through the evidence theory (late fusion).

An overview of the proposed framework is presented in the following section.

2. Overview of the proposed framework

The proposed framework, illustrated in Fig. 1, relies on the Bag of Visual Words (BoVW) model (Sivic and Zisserman, 2003) to characterize images. In the BoVW model, local image patterns are detected in images and each of these patterns is associated with a *visual word*. Each image, or each region of an image, is then associated with a *bag of visual words*, containing all visual words extracted from that image or region. Next, that image or region can be characterized by a histogram of visual words contained in the associated bag.

The first thing to define in the BoVW model is a dictionary of visual words. In a data mining context, it would be useful to work with more or less precise visual words. For instance, to extract rules about lesions in general, a ‘lesion’ visual word would be useful. But other rules may focus on a particular subtype of lesions, such as bright lesions, so a ‘bright lesion’ visual word would also be useful. Some rules may even focus on a particular lesion, such as exudates, so an ‘exudate’ visual word would be useful as well, etc. Therefore, unlike the usual BOVW model, a multigranular visual word dictionary is designed, as presented in Section 3.

The second thing is to put visual words in context. Experts may interpret an image pattern differently depending on its location with respect to the main retinal landmarks: the optic disk, the fovea and the blood vessels in particular. For instance, lesions in the fovea likely affect the patient’s sight more severely than the exact same lesions in the periphery of the retina. So, in order to reliably interpret a pattern within a fundus photograph, this pattern needs to be localized with respect to retinal landmarks. This is done by grouping visual words into bags, through meaningful region definitions, with respect to these landmarks. Some pathologies may affect the macula as a whole, others may only affect the fovea or the perifovea. So, various region sizes should be defined: each retina is divided into a cascade of regions. Each region is described by one histogram of visual words per level of lexical