Expert Systems With Applications 100 (2018) 224–233

Contents lists available at ScienceDirect

Expert Systems With Applications

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

A sequential search-space shrinking using CNN transfer learning and a Radon projection pool for medical image retrieval

Amin Khatami[a,*], Morteza Babaie[b,c], H.R. Tizhoosh[c], Abbas Khosravi[a], Thanh Nguyen[a], Saeid Nahavandi[a]

[a] Institute for Intelligent Systems Research and Innovation (IISRI), Deakin University, Australia
[b] Department of Mathematics and Computer Science, Amirkabir University of Technology, Iran
[c] KIMIA Lab at the University of Waterloo, Canada

ARTICLE INFO

Article history:
Received 29 July 2017
Revised 10 January 2018
Accepted 31 January 2018
Available online 6 February 2018

Keywords:
Content-based image retrieval
CBIR
Medical imaging
Deep learning
Radon

ABSTRACT

Closing the semantic gap in medical image analysis is critical. Access to large-scale datasets might help to narrow the gap. However, large and balanced datasets may not always be available. On the other side, retrieving similar images from an archive is a valuable task to facilitate better diagnosis. In this work, we concentrate on forming a search space, consisting of the most similar images for a given query, to be used for a similarity-based search technique in a retrieval system. We propose a two-step hierarchical shrinking search space when local binary patterns are used. Transfer learning via convolutional neural networks is utilized for the first stage of search space shrinking, followed by creating a selection pool using Radon transform for further reduction. The difference between two orthogonal Radon projections is considered in the selection pool to extract more information. The IRMA dataset, from ImageCLEF initiative, containing 14,400 X-ray images, is used to validate the proposed scheme. We report a total IRMA error of 168.05 (or 90.30% accuracy) which is the best result compared with existing methods in the literature for this dataset when real-time processing is considered.

1. Introduction

The main challenge in image search is to find the most relevant information among a set of images for a given query image. In medical domain, searching for similar cases in terms of the same anatomy and/or pathology can serve as “virtual peer review” for diagnostic purposes. Retrieving similar cases along with associated information and reports from repositories when the query is being treated by general practitioners and radiologists can establish a new level of comparative diagnostic which is presently absent and can immensely contribute to more accurate and robust diagnosis (Khatami et al., 2017d; 2017e; Zhang, Song, Cai, Liu, Liu, Pujol, Kikinis, Xia, Fulham, & Feng, 2016; Zhang, Huang, Li, & Metaxas, 2012).

Nowadays, the number of digital images generated by medical imaging devices is raising enormously. Managing and analyzing these large-scale repositories are becoming significantly complicated. If the size of the search space is too large making the image information not retrievable, the *semantic gap* will remain as an insurmountable challenge (Smeulders, Worring, Santini, Gupta, & Jain, 2000). The semantic gap is the amount of image information lost to a numerical representation of some features.

To make content-based image retrieval (CBIR) feasible, a robust searching scheme to find similar cases in large archives is needed. To overcome this challenge, hand-crafted features, deep features, dictionary approach, and other algorithms are used (Avni, Greenspan, Konen, Sharon, & Goldberger, 2011; Greenspan & Pinhas, 2007; Khatami, Khosravi, Lim, & Nahavandi, 2016; Khatami, Khosravi, Nguyen, Lim, & Nahavandi, 2017b; Kumar, Kim, Cai, Fulham, & Feng, 2013; Nanni, Brahnam, & Lumini, 2010; Rahman, Bhattacharya, & Desai, 2007; Xu, Lee, Antani, & Long, 2008).

There are two distinct strategies for image retrieval in a CBIR system. The first one is retrieving particular organs in particular modalities such as retrieving malignant lung nodules (Khatami et al., 2017c; Pan, Qiang, Yuan, & Wu, 2016) and liver lesions in CT Images (Napel et al., 2010), and chest structures from X-ray images (Shin et al., 2016). The second strategy concentrates on global similarity search in heterogeneous archives to identify and

* Corresponding author.
E-mail addresses: amin.khatami@deakin.edu.au, skhatami@deakin.edu.au (A. Khatami), mbabaie@uwaterloo.ca (M. Babaie), tizhoosh@uwaterloo.ca (H.R. Tizhoosh), abbas.khosravi@deakin.edu.au (A. Khosravi), thanh.nguyen@deakin.edu.au (T. Nguyen), saeid.nahavandi@deakin.edu.au (S. Nahavandi).

https://doi.org/10.1016/j.eswa.2018.01.056
0957-4174/

retrieve similar cases (Baâzaoui, Barhoumi, Ahmed, & Zagrouba, 2018; Greenspan & Pinhas, 2007; Seo, 2007). We follow the second strategy to propose a robust CBIR system which is based on two sequential procedures for shrinking the search space using deep networks and local similarity-based search techniques. However, utilizing deep networks for image retrieval in medicine is a challenging task due to two factors: (1) lack of large and "labelled" datasets for training, and (2) naturally existence of case "imbalance" in the medical domain.

Applying robust similarity search on properly shrunk search space, which is expected to make the semantic gap smaller, is a key to design accurate retrieval systems. However, utilizing high-dimensional feature spaces and the choice of reliable distance metric should be taken into account in order to produce meaningful results (Aggarwal, Hinneburg, & Keim, 2001; Alonso & Contreras, 2016; Hinneburg, Aggarwal, & Keim, 2000; Müller, Michoux, Bandon, & Geissbuhler, 2004).

Motivated by achievements of deep learning in computer vision and the applicability and practicability of Radon transform in medical domain, we utilize transfer learning via a convolutional neural network (CNN), as well as a Radon-based selection pool to sequentially shrink the search space. The smaller the search space with several right candidates, the better performance achieved by using the descriptors such as local binary patterns (LBP).

The contributions of this research are started with operating on a strongly imbalanced dataset and delivering the best accuracy and performance reported in literature so far. As well, to the best of our knowledge, this is the first work which creates a feature vector based on a classification-based shrunk search space for further shrinking. This contribution significantly improves the performance, as seen later. We also propose a two-step sequential shrinking search space, using a CNN and Radon transform, resulting in improved retrieval accuracy.

The rest of the paper is organized as following: Literature review is presented in Section 2. Descriptions of the proposed model are presented in Section 3. Section 4 explains the experimental results along with the analysis and discussions. A comprehensive performance comparison is also reported in Section 4, followed by concluding remarks in Section 5.

2. Related works

Several essential stages should be followed to obtain the similarity among features, resulted in a robust CBIR system: (1) *Content description:* the features of color, shape, texture, and so on should be extracted from images, (2) *Feature vectors:* an integrated feature vector should be properly assembled describing the information of the query and the images in datasets. Note that efficiency is important in this stage, and (3) *Similarity measure:* the metrics calculating the similarity among the feature vectors is paramount.

Texture descriptors are commonly used in medical image retrieval systems (Junior, Delgado, Gonçalves, & Nunes, 2009; Vipparthi & Nagar, 2014). As shown in Kashif, Deserno, Haak, and Jonas (2016); Sargent, Chen, Tsai, Wang, and Koppel (2009), it seems that keypoint-based descriptors such as SIFT, SURF, and ORB are not able to generate reliable feature points for some types of medical images. Dense sampling methods such as LBP are well-known in retrieval domain. Apparently, a thorough investigation is often required to obtain the best and efficient feature vectors representing images (Babaie et al., 2017a; Brahnam, Jain, Nanni, Lumini et al., 2014; Pietikäinen, Hadid, Zhao, & Ahonen, 2011). Global features are also widely used in medical image retrieval (Kumar et al., 2013). Radon transformation is mostly a global descriptor which extracts information of images from different directions. This transformation is widely utilized in medical domain due to easy implementation and efficient matching (Babaie, Tizhoosh, Zhu, & Shiri, 2017b; Clack & Defrise, 1994; Metz & Pan, 1995; Weisi et al., 2011). Moreover, Radon-based features may result in an efficient retrieval system by creating a short-length feature vector, which is in contrast to the aforementioned descriptors (Tizhoosh, 2015).

Image retrieval in medical application (IRMA) benchmark, (Lehmann et al., 2005; 2004b), explained in experimental results section, is an quite interesting medical dataset which consists of X-ray images of different body parts for patients of different age and gender (see Fig. 1). The images are assigned *IRMA codes* for benchmarking. As mentioned in Tommasi, Caputo, Welter, Güld, and Deserno (2009), Khatami et al. (2017a), Khatami, Babaie, Khosravi, Tizhoosh, and Nahavandi (2018) and Liu, Tizhoosh, and Kofman (2016), many studies have been performed on this benchmark. Radon-based annotations for medical image retrieval were proposed byTizhoosh (2015), resulting in the IRMA error of 470.57 (3) on the total test set of IRMA benchmark. In another study developed in Tizhoosh, Zhu, Lo, Chaudhari, and V. (2016a), a small number of equidistant projections of Radon was examined to generate a retrieval system on a set of IRMA dataset.

Based on a wealth research reports, a combination of Radon transformation with deep learning techniques on IRMA benchmark has proven to be a reliable approach to image search (Liu et al., 2016; Sze-To, Tizhoosh, & Wong, 2016; Tizhoosh, Mitcheltree, Zhu, & Dutta, 2016b).Liu et al. (2016) used CNN features (the information from the last fully connected layer) for a local search scheme, obtained by Radon barcodes to achieve the IRMA error of 224.13. Also, the IRMA error of 344.08 was obtained by Sze-To et al. (2016), using deep autoencoders and Radon projections. In another research, a deep auto-encoded Radon retrieval system was developed and achieved the IRMA error of 392.09 (Tizhoosh et al., 2016b).

As reported in Müller et al. (2009), an IRMA error of 178.93 was acquired by Idiap research team, utilising Support Vector Machines (SVMs) and the two descriptors of LBP, and modSIFT (Tommasi & Orabona, 2010). Avni, Goldberger, and Greenspan (2009) achieved an IRMA error of 169.5 by using a dictionary approach on IRMA dataset. More specifically, they developed a multi-resolution patch-based dictionary approach by using principal component analysis on the densely sampled patches, followed by an SVM classifier trained on the bag-of-words. Camlica, Tizhoosh, and Khalvati (2015) obtained an IRMA error of 146.55, which it is the lowest reported error so far. However, their saliency method is extremely sluggish such that they neglect the overhead for the saliency calculations and simply use the offline-generated maps during testing. For this reason, we do not compare our results with their method as this approach would be impractical for daily clinical practice.

3. Methodology

A sequential shrunk search space is introduced in this study for a uniform LBP descriptor (Ojala, Pietikainen, & Maenpaa, 2002). The shrunk space enables an efficient searching system. A distance metric is utilised to measure the similarity between two images. Properly shrinking the search space for LBP guarantees an accurate and robust retrieval system. The proposed retrieval system is summarized into three main parts, as depicted in Fig. 2. (1) First stage shrinking which is equipped by a transfer learning technique. A CNN model is utilised for this step. (2) Second stage shrinking which is obtained by defining a selection pool. Radon transformation is utilised to create the pool. (3) The last step is a local search procedure based on a similarity-based routine. The LBP descriptor is used to measure the similarity between images via the Manhattan metric.

A brief discussion of the two shrinking steps is explained, as follows.