Applied Soft Computing 30 (2015) 650–662

Contents lists available at ScienceDirect

Applied Soft Computing

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

Content-based image retrieval techniques for the analysis of dermatological lesions using particle swarm optimization technique

G. Wiselin Jiji [a,*], P. Johnson DuraiRaj [b]

[a] Department of Computer Science & Engineering, Dr. Sivanthi Aditanar College of Engineering, Tiruchendur 628215, India
[b] Dr. Sivanthi Aditanar College of Engineering, Tiruchendur 628215, India

ARTICLE INFO

Article history:
Received 4 August 2014
Received in revised form 25 January 2015
Accepted 26 January 2015
Available online 7 February 2015

Keywords:
Retrieval
Features
Classification
Particle swarm optimization

ABSTRACT

This method presents extraction of effective color and shape features for the analysis of dermatology images. We employ three phases of operation in order to perform efficient retrieval of images of skin lesions. Our proposed algorithm used color and shape feature vectors and the features are normalized using Min–Max normalization. Particle swarm optimization (PSO) technique for multi-class classification is used to converge the search space more efficiently. The results using receiver operating characteristic (ROC) curve proved that the proposed architecture is highly contributed to computer-aided diagnosis of skin lesions. Experiments on a set of 1450 images yielded a specificity of 98.22% and a sensitivity of 94%. Our empirical evaluation has a superior retrieval and diagnosis performance when compared to the performance of other works. We present explicit combinations of feature vectors corresponding to healthy and lesion skin.

1. Introduction

Content-based image retrieval (CBIR) is an important tool for handling large amounts of visual information in medical applications [1,2]. Three key steps can be identified in a generic CBIR system: (a) the extraction of valid information from images, (b) the computation of a suitable distance measure which quantifies the similarity of a query image from the reference images, and (c) the identification of a set of database images which display the smallest distances from given query. So an image query based retrieval system retrieves the most similar images. Napel et al. [3] developed a retrieval system to facilitate the retrieval of radiologic images that contain similar-appearing lesions with a database of computed tomography (CT) images of the liver. They have computed image features on the basis of image texture and boundary sharpness and these features are weighted. An independent reference standard was created for pair wise image similarity. This was used in a leave-one-out cross-validation to train weights that optimized the rankings of images in the database in terms of similarity to query images.

Potential descriptors of image content include features related to color, shape, texture or spatial information. Color and shape have proved as efficient descriptor in dermatology. Celebi et al. [4] presented a systematic overview of recent border detection methods: clustering followed by active contours are the most popular. Numerous features have been extracted from skin images, including shape, color, texture and border properties [5]. A comprehensive review of CBIR in medical imaging is presented in [6]. Its scope suggested by various researchers further set ups a roadmap for research in the field of CBIR system for medical image database. Classification methods used to index the pigmented skin lesion and dermoscopy database images are range from discriminate analysis to neural networks and support vector machines [7,8]. For the analysis of pigmented lesions of dermoscopic images are presented by Sáez et al. [9]. To improve the diagnosis accuracy they have presented algorithmic method and also a new 7-point checklist method. This is a simplification pattern analysis. It classifies seven features related with local patterns. In CBIR system, retrieval of images made simple by index the similar images by few clustering techniques. Relevance feedback, log-based clustering, hierarchical clustering and retrieval dictionary based clustering are the key methods used in the earlier works which were surveyed in [10]. These methods are mainly developed for images acquired by epiluminescence microscopy (ELM or dermoscopy) and they focus on melanoma, which is actually a rather rare, but quite dangerous, condition whereas other skin cancers are much more common. In

* Corresponding author. Tel.: +91 04639 242482; fax: +91 04639 243188; mobile: +91 9443087064.
E-mail address: jijivevin@yahoo.co.in (G.W. Jiji).

http://dx.doi.org/10.1016/j.asoc.2015.01.058
1568-4946/

[11], Chun et al. proposed content-based image retrieval method based on an efficient combination of multi-resolution color and texture features. Experimental results show that the proposed method yields higher retrieval accuracy than some conventional methods even though its feature vector dimension is not higher than those of the latter for six test DBs. Also, they demonstrate more excellent retrieval accuracy for queries and target images of various resolutions. In [12], CBIR method has been proposed which uses the combination of dominant color, GLCM texture and gradient vector flow field representation of shape. A total of 39 features covering color, texture and shape proved that their proposed method yielded higher average precision and average recall. In addition, that method almost always showed performance gain of average retrieval time over the other methods.

As conventional color features used in CBIR, there are color histogram, color correlogram, and dominant color descriptor (DCD). Color histogram is the most commonly used color representation, but it does not include any spatial information. Color correlogram describes the probability of finding color pairs at a fixed pixel distance and provides spatial information. Therefore color correlogram yields better retrieval accuracy in comparison to color histogram. These low level visual features of the images are especially useful to represent and to compare images automatically. In the concrete selection of color, texture and shape description [13], they use dominant colors, gray-level co-occurrence matrix and gradient vector flow field. Rahman et al. [14] presented a CBIR system for dermatoscopic images. Their approach includes image processing, segmentation, feature extraction (color and textures) and similarity matching. The extracted features are reproduced as PCA sub-space for similarity matching. Experiments on 358 images of pigmented skin lesions from three categories (benign, dysplastic nevi and melanoma) are performed. A quantitative evaluation based on the precision curve shows the effectiveness of their system to retrieve visually similar lesions.

CBIR of medical images is a useful tool, and could provide physicians with assistance in the form of a display of relevant past cases with proven pathology, along with the associated clinical, diagnostic, and other information. The appearance of a wound or lesion provides important clues that can help with the diagnosis, determination of severity, and the prognosis of healing. The proposed system is different from the work discussed in literature on the following grounds:

- The work concentrated on 125 different skin lesions.
- Assist the physician in diagnosis along with symptoms, treatment plan and case history.
- A novel approach PSO for multi-class classification is used.

In this work, experimental dataset is divided into two parts: developmental data set which is used as an image library and an unlabeled independent test dataset. We employ and compared two different methods to learn favorable feature representations for this special application.

2. Methodology

A practical approach to the retrieval of dermoscopy images is presented. CBIR undertakes the extraction of feature descriptors from each image, which, consequently, is used for computing similarity between images during the retrieval procedure. Fig. 1 shows an overview of the proposed approach. In the first phase, preprocessing operation is carried out to select the region of interest. In the second phase, feature extraction for the extracted region and feature selection are performed. In the third phase, indexing and retrieval operation would be achieved. Initially 19 shape and 10 color feature vectors are used extracted from skin images and applying dimensionality reduction using B4 method, and we selected the most promising 5 shape and 5 color feature vectors using 'scree graph'. Finally, we have implemented a fusion strategy color and shape feature vectors for image retrieval system using particle swarm optimization for multi-class classification.

2.1. Dataset description

The digital dermoscopy images and its symptoms, treatment plans and case histories were collected from Government Medical College, Tirunelveli, India, and some from online repositories (http://library.med.utah.edu, http://www.medicinenet.com). These were annotated by Dr. Sathish Kumar, Professional Dermatologist. We have collected data of 125 diseases with 1450 color images in resolution of 768×512 pixels and we experimented 320 images which belong to 10 classes. The classes examined in this work are anthrax, boils, burns, acne, impetigo, itching, keloid, psoriasis, ringworm, and skincancer.

2.2. Border detection

The first step in the computerized analysis of skin lesion images is the detection of the lesion borders. The importance of the border detection for the analysis is twofold. Firstly, the border structure provides vital information for accurate diagnosis. Many clinical features such as asymmetry and border irregularity are calculated

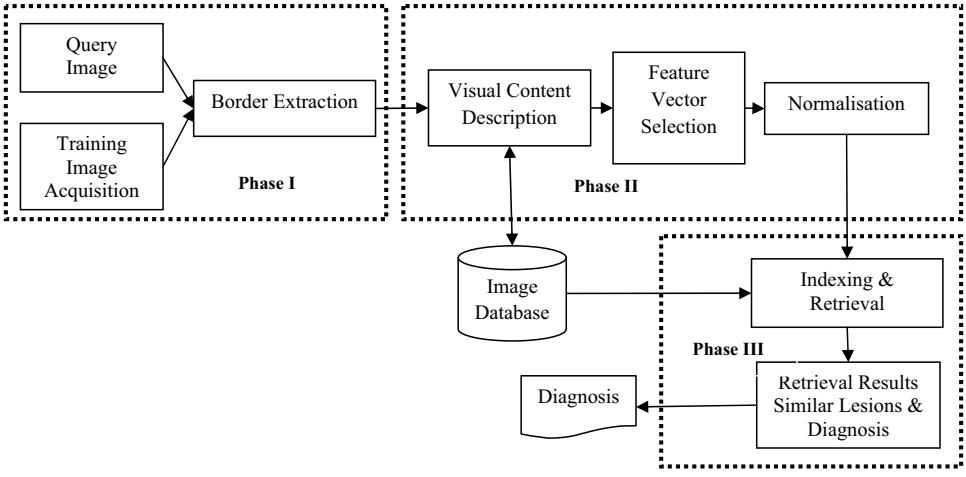

Fig. 1. Proposed architecture.