Discrete Applied Mathematics 197 (2015) 103–113

Contents lists available at ScienceDirect

Discrete Applied Mathematics

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

A new measure for comparing biomedical regions of interest in segmentation of digital images

Aura Conci [a,*], Stephenson S.L. Galvão [a], Giomar O. Sequeiros [a], Débora C.M. Saade [a], Trueman MacHenry [b]

[a] Computer Institute, Fluminense Federal University, Brazil

[b] Department of Mathematics and Statistics, York University, Canada

ARTICLE INFO

Article history:
Received 5 January 2014
Received in revised form 30 May 2015
Accepted 4 June 2015
Available online 10 July 2015

Keywords:
Biomedical image processing
Image segmentation
Infrared imaging
Hausdorff distance
Discrete sets

ABSTRACT

The segmentation of the region of interest (ROI) of digital images is generally the first step in the pattern recognition (PR) procedure. Automatic segmentation of biomedical images is desirable and comparisons among new approaches, by using available databases, are important. We present a new approach to compute the Hausdorff distance (HD) between digital images. Although HD is the most used distance estimator among sets, we show why it is not suitable for biomedical applications. In this paper, a new technique to define the degree of correction of the ROI is developed to serve as a basis for the comparisons used to validate works on segmentation of biomedical images. As for online diagnosis, the comparison among possible techniques must be efficient enough to: (1) be done in real time (i.e. during the examination), (2) allow the inclusion of priority aspects, and (3) be intuitive and simple enough to be easily followed by people with no computational or mathematical background. We develop a new index by considering the expectations of the medical doctors who are using computer systems for diagnostic aids, and take into consideration how these systems use ROIs to extract feature properties from the examinations. We discuss conditions for empirically defining a measure for calculating similarities and differences between ROIs. The proposed method is applied to both real and simulated data examples.

1. Introducing the problem

Dedicated digital image processing chips convert raw data acquired from digital cameras into a discrete set of points. The quality of the digital images is related to the numbers of discrete points used (named *pixels* in computer graphics and in image processing) and to the resolution used by the digital camera. Image processing is typically a special application of methods from discrete mathematics [9]; it considers several tools to manipulate the images in many ways. In this work we are interested in a particular type of digital camera that can acquire the infrared (IR) frequency of each discrete point in a scene (Fig. 1(a)). Even moreparticularly, we are interested in the relationship with this IR frequency and the temperature of the human body, when such an element is included in the frame (Fig. 1(b)). We consider that, for each discrete point that is captured, we have its coordinates as well as its temperature, which is a real number representing a color and able to take a finite number of different values (Fig. 1(a) and (c)). These temperatures of the human body can be used for diagnosis, where

* Corresponding author.
E-mail addresses: aconci@ic.uff.br (A. Conci), sgalvao@ifpi.edu.br (S.S.L. Galvão), goliver@ic.uff.br (G.O. Sequeiros), debora@ic.uff.br (D.C.M. Saade), machenry@mathstat.yorku.ca (T. MacHenry).

http://dx.doi.org/10.1016/j.dam.2015.06.002

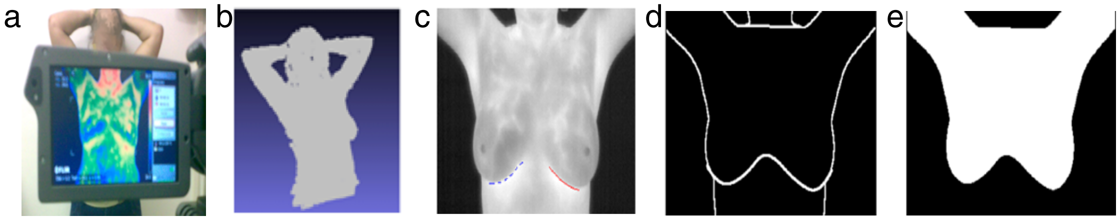

Fig. 1. IR camera and the captured scene, where the temperature of each discrete point is represented by colors (a), each discrete point of the image (b), representation of the temperature using gray level and the first step of segmentation process (c), that is, the inframammary folds of the right and left breast of the patient defined by dashed and continuous lines respectively. Two possible ways to define the ROI are by its boundaries (d) and, by labeling the pixels belonging to ROI (e).

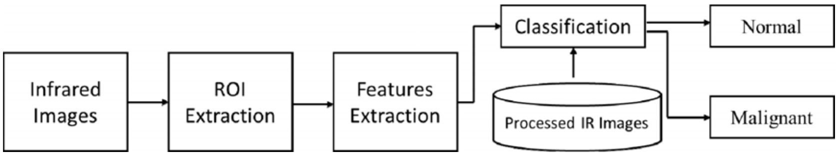

Fig. 2. Steps of IR based CADs for breast diseases.

the temperatures are restricted to a special region of the body. Proper definition of such a region is very important for the following steps of the application. When algorithms define these boundaries, the distances among the correct borders, and those achieved by a given implementation, need to be characterized. The Hausdorff distance (HD) is a metric for evaluating the distance between discrete sets of points [2]. This paper presents a new algorithm for HD computation for discrete images, based on the geometry of parallel bodies [5], and the characterization of the distance among sets of points [11].

Pattern Recognition (PR) and image processing techniques have been applied in analysis of the most common types of medical diagnosis [16]. In these techniques, the first step after acquiring the image (Fig. 2) is to separate the important elements, that is, to obtain the region of interest (ROI) [19]. Two common representations of the ROI are by its boundary or by its area. A contrasting line can be used in the graphical representation of ROI in the first case (Fig. 1(d)). In the second case, the ROI can be identified by labeling its pixels as "1" or white (Fig. 1(e)). In the IR frontal breast examination, right and left orientation always refers to the patient. Thus, in Fig. 1(c), a blue line defines the lower boundary (inframammary fold) of right side of the patient, and a red line defines the lower boundary of the left breast.

Typically, there are two ways to design image segmentation algorithms, with user interactions or automatically. In the first case, there is some degree of dialog with the user, for instance, in the identification of the important objects in the image. In the second case, the segmentation is realized without any human intervention. In most of the cases of computer aided diagnosis (CAD), a complete automatic procedure of ROI segmentation must be carried out [7]. That is, this region must be found with no user interaction. However, due to the complexity and importance of proper ROI definition for the next steps of these CAD systems (that are represented in Fig. 2), the exactitude of ROI identification algorithms must be proved. To evaluate the ROI correctness, a large number of cases, which are segmented by a group of experts, needs to be considered. However, it is not easy to find and persuade good specialists to spend their time doing this type of work. In addition, more than one result for each case and image is needed in order to achieve segmentations free of personal bias. These aspects make the set of segmentations considered to be correct of great value in the research. They constitute the commonly named "truest sets" for comparisons of results, or what is called the "ground truth" (GT) [4,6].

After a number of automatic segmentations have passed through the comparison stage with the GT (composed by a combination of correct segmentations), they are included in CAD systems. Moreover, these systems, in most cases, perform segmentation algorithms in real time, that is, during the patient examination period. In this way, in case of doubt, there is the possibility of repeating the examination process. In order to verify the adequacy of different approaches, numerical (rather than visual) techniques must be used to compare the ROIs. Keeping in mind all these aspects, it is clear why it is of great importance to have a metric to show how similar two ROI boundaries are. The definition of a fast, intuitive and correct evaluator for medical ROI comparisons is still an open problem. The presentation of a new index to compare ROIs for diagnostic systems is the goal of this work. It also presents a brief review of similarity measures currently used for ROI comparison. Moreover, it develops a new candidate for a measure with which to compare nonconvex, nondisjoint, but fully connected regions, that constitute the ROI of IR breast images, and discusses its advantages.

This work is based on an application under development in the Radiologic Service of the University Hospital of the Fluminense Federal University (UFF), Niteroi, Rio de Janeiro, Brazil: The project PROENG (http://visual.ic.uff.br/en/proeng/). The aim of this project is to research the possibilities of IR breast examinations to help breast disease detection, and to develop a methodology for machine learning and decision support system for the early detection of breast cancer and follow-up by adjuvant therapy.

The rest of the paper is organized as follows. The next section gives the reader some background about the importance of ROI validation. In Section 3, we consider details of our new index and other measure to compare segmentation approaches.