Chapter 11
A Probabilistic SVM Approach to Annotation of Calcification Mammograms

Chia-Hung Wei
Ching Yun University, Taiwan

Sherry Y. Chen
Brunel University, UK

ABSTRACT

Due to the increasing use of digital medical images, a need exists to develop an approach for automatic image annotation, which provides textual labels for images. Thus added labels can be used to access images using textual queries. Automatic image annotation can be separated into two individual tasks: feature extraction and image classification. In this paper, the authors present feature extraction methods for calcification mammograms. The resultant features, based on BI-RADS standards, make annotated image contents represent the correct medical meaning and tag correspondent terms. Furthermore, this paper also proposes a probabilistic SVM approach to image classification. Finally, the experimental results indicate that the probabilistic SVM approach to image annotation can achieve 79.5% in the average accuracy rate.

1. INTRODUCTION

In the last decade, a large number of digital medical images have been produced in hospitals. Such digital medical images include X-ray, computed tomography (CT), magnetic resonance imaging (MRI), functional magnetic resonance imaging (fMRI), magnetic resonance spectroscopy (MRS), magnetic source imaging (MSI), digital subtraction angiography (DSA), positron emission tomography (PET), ultrasound (US), nuclear medical imaging, endoscopy, microscopy, scanning laser ophthalmoscopy (SLO), and so on. These medical images are stored in large-scale image databases.
and can facilitate medical doctors, professionals, researchers, and medical college students to compare the current patients and the previous cases, and provide valuable information for their studies. Due to the increasing use of digital medical images, there is a need to develop advanced information retrieval techniques. Among various information retrieval techniques, automatic image annotation is considered as a prerequisite task for image database management (Hersh, 2009). Automatic annotation can provide textual labels for images and thus added labels can be used to access images using textual queries, thereby improving the effectiveness of browsing and searching of large medical image databases.

Automatic image annotation has been a hot topic in the areas of multimedia, information retrieval, and machine learning. To correspond to this trend, this paper presents an image annotation scheme, which includes mammographic feature extraction and an approach to do automatic mammogram annotation. The rest of this paper is organized as follows: Section 2 reviews methods of visual feature extraction and automatic image annotation in the medical domain. Section 3 provides a set of features for describing calcification lesions in mammogram. Section 4 proposes a probabilistic SVM approach to image classification. The experimental results are presented and discussed in Section 5. Finally, the conclusion is made in Section 6.

2. LITERATURE REVIEW

2.1 Representation of Medical Image Content

Automatic image annotation refers to a technique that automatically assigns a set of linguistic terms to images in order to categorize the images conceptually and provide means for effectively accessing images from databases (Deselaers, Deserno, & Muller, 2007). To make computers automatically assign linguistic terms to images, the region of interests in images need to be represented from corresponding visual features. Visual features, also called low-level, are objectively derived from the images, rather than referring to any external semantics (Feng, Siu, & Zhang, 2003a). As the visual features extracted from the images should be meaningful for image seekers, the visual features used in the image retrieval systems are mainly divided into three groups: color, shape, and texture.

Color, one of the most frequently used visual features for content-based image retrieval, is considered as a powerful descriptor that simplifies object identification (Gonzalez & Woods, 2002). Several color descriptors have been developed from various representation schemes, such as color histograms (Ouyang & Tan, 2002), color moments (Yu, Li, Zhang, & Feng, 2002), color edge (Gevers & Stokman, 2003), color texture (Guan & Wada, 2002), and color correlograms (Moghaddam, Khajoie, & Rouhi, 2003). For example, color histogram, which represents the distribution of the number of pixels for each quantized color bin, is an effective representation of the color content of an image. The color histogram can not only easily characterize the global and regional distribution of colors in an image, but also be invariant to rotation about the view axis.

Color allows images to reveal many lesion characteristics, which is useful to retrieve medical images, (Tamai, 1999). Color also plays an important role in morphological diagnosis (Nishibori, Tsumura, & Miyake, 2004). Color medical images are usually produced in different departments and by various devices. For example, color endoscopic images are taken by a camera that is put into the hollow organs of the body, such as stomachs and lungs. A common characteristic of such images is that most colors are made of various stains, though fine variations of natural colors are crucial for diagnosis. Nishibori (Nishibori, 2000) pointed out that problems in color medical images include inaccurate color reproduction, rough gradations of color, and insufficient density of pixels. Therefore,