Multimodal Information Integration and Fusion for Histology Image Classification

Tao Meng, University of Miami, USA
Mei-Ling Shyu, University of Miami, USA
Lin Lin, University of Miami, USA

ABSTRACT

Biomedical imaging technology has become an important tool for medical research and clinical practice. A large amount of imaging data is generated and collected every day. Managing and analyzing these data sets require the corresponding development of the computer based algorithms for automatic processing. Histology image classification is one of the important tasks in the bio-image informatics field and has broad applications in phenotype description and disease diagnosis. This study proposes a novel framework of histology image classification. The original images are first divided into several blocks and a set of visual features is extracted for each block. An array of C-RSPM (Collateral Representative Subspace Projection Modeling) models is then built that each model is based on one block from the same location in original images. Finally, the C-value Enhanced Majority Voting (CEWMV) algorithm is developed to derive the final classification label for each testing image. To evaluate this framework, the authors compare its performance with several well-known classifiers using the benchmark data available from IICBU data repository. The results demonstrate that this framework achieves promising performance and performs significantly better than other classifiers in the comparison.

Keywords: Bio-Image Informatics, Biomedical Imaging Technology, C-Value Enhanced Weighted Majority Voting (CEWMV), Collateral Representative Subspace Projection Modeling (C-RSPM), Histology Image Classification

1. INTRODUCTION

The bio-imaging techniques such as fluorescent microscopy, two-photon-laser scanning microscopy, and electron microscopy have become essential tools for exploring the structures and functions of organisms. With the advancement in the automatic imaging technologies, there is a rapid increase of biomedical imaging data sets in recent years, ranging from X-ray, CT for disease diagnosis, to in situ hybridization (ISH) imaging for analyzing gene expression patterns. The number of bio-images is increasing on a scale comparable to that of the genomic revolution (Hamilton et al., 2006).
The huge amounts of biomedical image datasets present significant challenges for traditional analysis methods based on manual annotation and human labeling. Therefore, utilizing computing technologies in automatic image processing and analyses has become popular research topics. With the collaboration of biomedical scientists and computer scientists, the area of “bio-image” informatics (Peng, 2008), which is a new branch of bioinformatics, was developed and a huge amount of computer-based algorithms for managing, indexing, and analyzing bio-image data sets has been introduced. Examples of such applications include automatic cell detection systems (Long, Cleveland, & Yao, 2010; Huang, Sun, & Hu, 2009), bio-image segmentation systems (Bae, Pan, Wu, & Badea, 2009; Madhloom, Kareem, Ariffin, & Zaidan, 2010), cell phenotype classification systems (Minamikawa et al., 2003), etc. A good review of this area was given in (Peng, 2008).

Histology is an essential tool in biomedical research field to examine microscopical anatomy of cells and tissues of plants and animals in order to infer the functional semantics of organisms. In one assay of histology experiment, the tissues or cells are stained so that the structure could be examined by human experts for annotating structural characteristics. With the deluge of the histology data sets, computer-based histology image analysis, which saves human labors and decreases the inter-intra variance, is demanded for automatic management and analyses of histology data and databases.

One of the critical tasks in this area is to classify the raw histology images based on the phenotype. By classifying images into different categories, it not only helps the medical scientists to make comparison within and cross varieties but also facilitates the computer scientists to build efficient indexing and retrieval systems. As a matter of fact, the diagnosis procedure is a binary classification problem itself. However, the task is quite challenging in this scenario for three reasons. First, histology images are non-stationary images. In other words, each region of the raw image could have distinct characteristics. Second, the variance in operating conditions in the laboratory increases the effect of artifacts, which leads to the increase of the intra-class differences. Third, the inter-class difference is relatively small and it is difficult for human to distinguish. In comparison among performances for different benchmark data sets in (Huang, Sun, & Hu, 2009), it showed that the accuracies of three histology data sets, Liver Aging, Liver Gender, and Lymphoma were relatively lower than those of others, which demonstrated the challenge of classifying these data sets.

There are some applications developed for analyzing histology data sets. The current state-of-art technologies can be summarized into three categories. The first category is developed for general biological image classifications. In order to accommodate to heterogeneous characteristics for different data sets, these applications usually encompass a large pool of features and carry the burden of huge computational complexity for feature extraction and feature selection. One example is the Wndchrm (Shamir et al., 2008) developed in the Laboratory of Genetics in National Institute of Aging. It contained 2659 features from raw pixels and transforms of images, such as Gabor Filters and Chebyshev-Fourier features, and utilized the nearest neighbor algorithm to perform the classification. CellProfiler (Carpenter et al., 2006) and EnhancedCellClassifier (Misselwitz et al., 2010) applied the support vector machine with radial basis function (RBF) kernel as the classifier and demonstrated relatively good performance especially on cell images.

The other category of research endeavor lies in developing specific applications for one or few data sets. In order to perform these tasks, many researchers in this domain bring techniques from image processing and pattern recognition fields. In the field of pattern recognition, transforming the images into good feature measurements is crucial for success. Therefore, a lot of approaches have been proposed to extract data-dependent features for a certain histology data set. A typical work of this type was presented in (Qureshi, Raipoot,
Requirements to a Search Engine for Semantic Multimedia Content
Lydia Weiland, Felix Hanser and Ansgar Scherp (2014). *International Journal of Multimedia Data Engineering and Management* (pp. 53-65).
[www.igi-global.com/article/requirements-to-a-search-engine-for-semantic-multimedia-content/120126?camid=4v1a](www.igi-global.com/article/requirements-to-a-search-engine-for-semantic-multimedia-content/120126?camid=4v1a)

Music Control in an Interactive Conducting System Using Kinect
[www.igi-global.com/article/music-control-in-an-interactive-conducting-system-using-kinect/103010?camid=4v1a](www.igi-global.com/article/music-control-in-an-interactive-conducting-system-using-kinect/103010?camid=4v1a)