Chapter 1

Introduction

Computational power has dramatically increased in the past decades which has led to developments in the area of digital microscopy. It has introduced more systems to automate tissue processing, staining, image acquisition and management of virtual slides. These developments make microscopic studies more amenable for use of computer aided image analysis, pattern recognition and machine learning techniques.

Currently, most of the pathological diagnosis is done based on evaluation by pathologists which is subjective and qualitative. Researchers from both computer science and medical science recognize the importance of quantitative, image-based evaluation of microscopic slides. It is important not only for diagnostic purposes but also for gaining insight into the disease process. Computer Aided Diagnosis (CAD) plays a very important role in increasing the performance of a human expert by providing quantitative information which can be complementary to the judgment of the expert. CAD also reduces the workload of a medical professional by filtering out obvious cases from more difficult suspicious cases for manual evaluation.

1.1 Image Analysis

Image analysis has become a powerful scientific tool and offers a whole range of measurements of optical parameters and color characteristics of objects. It allows obtaining
more precise and detailed data, thus enabling better comparison of different objects. Identifying particular objects, counting and measuring their size, shape, position, density and other similar properties, are tasks that are well within the power of computers. These tasks can be done relatively quickly by a computer with excellent reproducibility. Visual interpretation of images is an integral part of medical diagnostic procedures. However, it is tedious, time consuming and error prone. Computerized image analysis has been widely used in medical research facilitating quick, repeatable and objective evaluation, thus making it possible to obtain statistically significant results.

**Medical Image Analysis**

Medical image analysis is a specialized area that aims to extract diagnostically important interpretation and information from images in a repeatable and objective manner. Medical images can be classified into two types based on their origin. Radiological images, which are acquired through medical imaging modalities like Computed Tomography, Magnetic Resonance Imaging etc., are of the first type whereas images from medical microscopy are classified as the second type. Radiological images are generally grayscale images. There has been a lot of research on analysis of radiological images. This study focuses on microscopy images, which often are color images.

Observations based on microscopy images of histopathological specimen are used in screening, diagnosis, prognosis, therapy and in medical research. Decision on most of the severe pathologies is confirmed only after microscopical examination of the affected cells or tissues. Microscopical examination is carried out to obtain quantitative features like Deoxyribo Nucleic Acid(DNA) content, cell/nuclei size, cellular density, labeling intensity, count of cells, analysis of tissue architecture, amount of a specific cell component and to understand biological mechanisms such as cell proliferation and differential genetic expression. Quantitative analysis is performed on microscopy images to extract various kinds of measurements and to evaluate their statistical significance. Broadly, four techniques are used for microscopic image analysis. They are histologi-
Histology is the study of microscopic anatomy of cells and tissues of plants and animals. It is performed by examining a thin slice of tissue under a light microscope. For this, the appropriate tissue is collected and fixed. The purpose of fixation is to preserve tissues permanently in as life-like a state as possible. Tissue is processed into a form in which it can be made into thin microscopic sections. This is usually done by embedding with paraffin (Culling, 1974). Tissue embedded in paraffin can be sectioned at around 3 to 10 microns. To see the tissue under a microscope, the sections are stained with one or more pigments. The aim of staining is to reveal cellular components. Counter-stains are used to provide contrast. The staining process makes use of a variety of dyes that have been chosen for their ability to stain various cellular components of tissue. The stained section on the slide is covered with a thin piece of plastic or glass to protect the tissue from being scratched and to provide better optical quality for viewing under the microscope. The two main types of analyses in histological study of a specimen are histomorphologic analysis and quantification of substances. For histomorphologic study, area or number of cells are measured. Examples of such applications could be to count the number of positively stained nuclei, to measure the size of a cell, to assess fibrosis, cellular hypertrophy etc. Quantification of substances involves measurement of area and intensity for which histochemical, immunohistochemical or in-situ hybridization techniques are used. Diagnosis from a histological image remains the gold standard in diagnosing considerable number of diseases. Histological slides preserve the underlying tissue architecture and hence, can provide comprehensive view of disease and its effect on tissues.

Cytology is the study of structure, chemistry and function of cells. A large focus of microscopic image analysis has been on automation of cytological image analysis. Some of the quantitative features that can be obtained from cytological images are cell count, shape and size of nuclei, cell size and ratio of sizes of nuclei to cells.
Immunohistochemical analysis deals with localization of proteins in cells of a tissue section by exploiting the principle of antibodies binding specifically to antigens in biological tissues. It combines anatomical, biochemical and immunological techniques for identification of specific tissue components by means of a specific antigen/antibody reaction tagged with a visible label [Panja et al., 2007]. Immunohistochemical staining is widely used in the diagnosis of abnormal cells such as the cells found in cancerous tumors. Specific molecular markers are characteristic of particular cellular events such as proliferation or cell death. IHC is also widely used in medical research to understand the distribution and localization of biomarkers and differentially expressed proteins in different parts of a biological tissue. Thus, it has become a crucial technique and is widely used in both medical research laboratories as well as clinical diagnostics.

Histological, cytological and immunohistochemical analyses are based on light microscopy where as FISH analysis is based on flourescent microscopy.

1.2 Need for Automation

In the case of manual evaluation, the specimen is studied under a microscope and the pathologist/medical researcher makes judgment based on his/her experience. This judgment is qualitative, subjective and is prone to considerable variability. Manual evaluation requires prior training and the results are poorly reproducible. Errors are common even amongst experts. Productivity, accuracy and objective evaluation are the main challenges that the pathologists face. Pathology laboratories are under increasing pressure to process more slides and to complete more tests. Tackling huge amount of work manually places more demand on resources. Also, it is reported that after prolonged visual study, eye fatigue can significantly affect a person’s ability to discern differences in color shades. This can affect the accuracy of the evaluation. Also, the nature of human eye is such that every person sees an object slightly differently from the way others see it. Different observers may report seeing different features on the
same object, a single observer at different times may report seeing it differently. Thus, it is highly subjective.

Automation of image analysis has triggered a striking growth in CAD by enhancing the visual perceptive skills. In addition, the quantitative data thus generated may be subsequently correlated with other information to associate clinical status and tissue structure with genomic and proteomic levels. Great advantage of image analysis approach is quantification of the substance under study, in different tissues within single sample as well as among different samples, which in many cases cannot be detected with biochemical tools. Quantification of tissues is needed not only for a better understanding of physiological data, but also to establish correlations with elements seen histologically. An interesting result of digital image analysis for cancer detection is the potential to identify pre-malignant changes in tissues before any visual differences are perceptible. In the case of applications handling perishables, image analysis allows us to distinguish subtle differences that may be lost while using a more subjective analytical method. It also increases the productivity and hence can provide statistically significant results. Medical research involves study of large image datasets. Success in the research outcome depends heavily on accuracy, efficiency and consistency of evaluation. The need to draw accurate and reproducible conclusions is a fundamental objective in histological and cytological diagnosis. Automated image analysis can overcome the limitations of manual microscopy and thus improve the practice of pathology and also the quality of medical research.

1.3 Issues

Inconsistency of the study, anatomic inconsistency and staining inconsistency are the important factors to be considered as challenges to automation. Variations in tissue processing and section thickness can lead to wrong analysis. The automated slide mounting systems which are becoming more popular can handle these problems. Vari-
atation in the amount of stain used each time may also interfere with interpretation. This can be overcome by having a standard material being stained along with the substance under study. It is important to be reproducibly quantitative from day to day, between machines and between laboratories. Hence, complete color calibration is required. Multipoint calibration can be used for calibration of complete dynamic range which is necessary to ensure high quality results. Variation in background illumination is an important problem with microscopy images. Algorithms for illumination correction will have to be applied before any analysis can be performed on them. However, manual evaluation by an observer would be useful to manage artifacts, exceptional variations in the sample and for use of circumstantial knowledge.

1.4 Approaches

Traditional techniques for image analysis include a quite laborious semi-quantitative approach using a score ranging from 0 (no staining) to 4 (very intensive staining) for example, which is widely used by pathologists and researchers. In order to obtain more objective, reproducible, informative and easier measurements, digital image analysis techniques have been introduced.

Many research groups have put efforts to apply digital image analysis for the quantification of a biological substance for study of diseases like breast cancer (Caldwell et al., 2008; Aziz, 1992; Lehr et al., 1997; De Solorzano et al., 2002), lung diseases (De Boer et al., 2001), skin diseases (Herbin et al., 1990; Alla et al., 2011), nephropathy (Encarnacion et al., 2004), tuberculosis (Forero et al., 2004) and atheromatous lesions of human aorta and coronary arteries (Niendorf et al., 1990). Digital image analysis has also been tried for quantification of stained specimen for medical research in (Lehr et al., 1999; Pauschinger et al., 1999; Johansson et al., 2001; Maximova et al., 2006). Some groups performed analysis in grayscale, some others used threshold and area measurement. Some groups stored the possible color shades of region of interest in
Fig. 1.1: Images representing the collagen content represented in blue shades in control(A) and treated case(B)

a color file with the help of user interaction and the same file was consulted to evaluate rest of the images in the dataset. Color clustering algorithms (Masmoudi et al., 2009) and color quantization (Park et al., 2010) have also been tried to classify the regions in an image into foreground or background.

1.5 Motivation

Currently, most of the approaches to assess amount of stain and hence the biological substance under study, just determine the shades of a particular color of interest and select all pixels with those shades as foreground pixels. This approach does not take into account, the possibility of the substance being present in varying concentrations across the specimen. A technique which can grade the shades of the color of interest could be more useful to handle such variations. Considering that the biological sample is being used for study of growth of a particular kind of tissue, amount of tissue present in the sample needs to be determined. The stained image takes up different shades of color based on the density of tissue present. A certain amount of tissue if present densely packed in a small area would take up a strong color whereas, the same amount of tissue if spread in a larger area would take up paler shades.

Figure 1.1 shows the Masson’s Trichrome stained full thickness rat dermal wound images. Figure 1.1A represents the control and Figure 1.1B represents the $T/\beta_4$ treated case, used in the study on role of $T/\beta_4$ in dermal wound repair (Philp et al., 2003).
The amount of collagen content represented by shades of blue color in Figure 1.1B was estimated to be approximately double that of the control (Figure 1.1A), but this is highly subjective. In such cases, a scoring system for a color based on the depth of color is desirable. This facilitates a numeric measure for a quantity of the substance present, which otherwise would have to be described as grades. A technique which just considers the area would not be able to provide accurate results in these cases. Also, a small change in quantity of the substance, which cannot be visually perceived, can be observed through image analysis.

Manual evaluation introduces inter-observer and intra-observer variations. In the case of large image datasets, a fully automated analysis is highly desirable. If the algorithm can eliminate cases which can be classified with high confidence and filter them out, only the remaining cases need a review by a human expert. This kind of a decision support system can facilitate high throughput at the same time maintain accurate, repeatable and objective assessment.

1.6 Objectives and Scope

The objective of this work is to develop a new method of color representation so that the biological specimen, which takes up various colors on staining can be efficiently quantified. This kind of quantification can be useful for medical research and diagnostic applications. The goals of this research are summarized as follows.

1. To develop a new method to represent the similarity of color shades of pixels in an image to a reference color as a numerical measure.

2. To develop a color image analysis technique to quantify staining intensity in biological specimen based on this numerical measure.

3. To demonstrate the use of this representation for image analysis in medical research and diagnostic purposes.

The new method of color representation can be used to quantify the substance
under study, to represent the color texture of an image and also for accurate color segmentation. It is hoped that the proposed representation of color can help medical researchers to obtain accurate quantification of stained biological specimen and also to obtain accurate color segmentation.

1.7 Organization of the Thesis

This thesis comprises of seven chapters. This chapter provides an introduction to the thesis.

Chapter 2 provides a comprehensive review of literature in the related area. The approaches used by other research groups around the world in the related area and various image analysis tools used for this purpose are included in this review. Some studies have used general purpose image processing software with macros developed for a specific task. Such studies are also discussed here. In addition to this, studies on development of algorithm for specific analysis tasks are discussed. This survey attempts to bring out the motivation for this research work.

Chapter 3 describes the development of algorithm for staining intensity quantification and the image analysis tool using this algorithm named TissueQuant which could be used by a medical researcher. The algorithm is based on HSI color model and Gaussian weighting functions. TissueQuant was developed for analysis by a medical researcher as the end user of this tool. Hence, it has been specifically designed to be simple and suitable for a specific set of tasks involving quantification of specific substances or to filter out a particular color and its shades from the image under study. The features of TissueQuant, its calibration, validation and performance characterization are also described in this chapter. Comparison of performance of the proposed technique with the popular commercial software Adobe Photoshop is also provided.

Chapter 4 discusses quantification of a stained substance present in a biological specimen using the color scores. To demonstrate the use of color scores for quan-
tification in medical research, the study on evaluation of growth of osteoblasts and osteoclasts in bone section images is described. The efficacy of a plant extract as a drug to treat osteoporosis is evaluated by comparing its antiosteoporotic activity with that of the popular drug named Raloxifene which is currently being used for treatment of osteoporosis.

Chapter 5 discusses the development of computer aided diagnostic systems using color scores. Two applications of CAD are described here. In the first CAD system, the color scores are used for image analysis of immunohistochemically stained samples of breast biopsies for diagnosis of breast cancer. Moreover, investigation on inter-observer and intra-observer variations in grading breast cancer is carried out to highlight the benefit of using CAD system for this purpose. This chapter also describes the development of a CAD system for diagnosis of malaria using color scores for segmentation of specific components of blood cells in a thin blood smear image. Description of implementation of a prototype of an android application for use as a telemedicine system is also provided in this chapter.

Chapter 6 describes the use of TissueQuant software for morphometric analysis of nerves of forearm. In this application, the color segmentation is based on the color scores obtained through the TissueQuant algorithm. Nerve cross sectional area, fascicular area, number of fascicles, amount of adipose tissue, number of sympathetic fibers and their areas are the morphometric parameters measured.

Chapter 7 presents a summary of the research work and highlights the contributions of this work. It discusses the benefits of this work in the area of medical research and diagnostic applications. This chapter also suggests directions for further research.

The next chapter provides a detailed survey of state of the art literature in the related area.