A STUDY OF X-RAY IMAGE PERCEPTION FOR PNEUMOCONIOSIS DETECTION

Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science (by Research)
in
Computer Science and Engineering

by

Varun Jampani
200502027
varunjampani@research.iiit.ac.in

Center for Visual Information Technology
International Institute of Information Technology
Hyderabad - 500 032, INDIA
January 2013

International Institute of Information Technology
Hyderabad, India

CERTIFICATE

It is certified that the work contained in this thesis, titled 'A study of X-ray Image Perception for Pneumoconiosis Detection' by VARUN JAMPANI, has been carried out under my supervision and is not submitted elsewhere for a degree.

Date

Adviser: Prof. Jayanthi Sivaswamy

To my *grandfather* K. Anjaneyulu

Acknowledgments

I sincerely thank my professor and thesis advisor Prof. Jayanthi Sivaswamy for her invaluable guidance and support in completing this thesis work. I also want to thank her for showing me how to be critical in doing research. Special thanks to my other professor Prof. Bipin Indurkhya who taught me the basics of doing research and how to be matured in our reasoning and thinking. Suggestions and advices from Dr. Vivek Vaidya were very helpful in experimental analysis. Thanks to him. Thanks to Prof. Rajeev Sangal and Prof. Kamal Karlapalem for backing my decision to take an year's leave during the course of my master's work. I like to acknowledge GE Global research for supporting major part of my research work.

I am highly indebted to my friends at IIIT-Hyderabad (Chandu, Rohit, Srujan, Srinivas, Raghudeep, Vamshi, Soni and Srikar) for always supporting and encouraging me. Special thanks to my friends at Dharamsala (Tenzin Bety, Tsering Bhuti and Phurbu Thackchoe) for being highly supportive in various aspects of my life. Finally, I want to thank my parents, grand parents and brother for being supportive and for giving me freedom in pursuing my interests.

Abstract

Pneumoconiosis is an occupational lung disease caused by the inhalation of industrial dust. Despite the increasing safety measures and better work place environments, pneumoconiosis is deemed to be the most common occupational disease in the developing countries like India and China. Screening and assessment of this disease is done through radiological observation of chest x-rays. Several studies have shown the significant inter and intra reader observer variation in the diagnosis of this disease, showing the complexity of the task and importance of the expertise in diagnosis.

The present study is aimed at understanding the perceptual and cognitive factors[1] affecting the reading of chest x-rays of pneumoconiosis patients. Understanding these factors helps in developing better image acquisition systems, better training regimen for radiologists and development of better computer aided diagnostic (CAD) systems. We used an eye tracking experiment to study the various factors affecting the assessment of this diffused lung disease. Specifically, we aimed at understanding the role of expertize, contralateral symmetric (CS) information present in chest x-rays on the diagnosis and the eye movements of the observers. We also studied the inter and intra observer fixation consistency along with the role of anatomical and bottom up saliency features in attracting the gaze of observers of different expertize levels, to get better insights into the effect of bottom up and top down visual saliency on the eye movements of observers.

The experiment is conducted in a room dedicated to eye tracking experiments. Participants consisting of novices (3), medical students (12), residents (4) and staff radiologists (4) were presented with good quality PA chest X-rays, and were asked to give profusion ratings for each of the 6 lung zones. Image set consisting of 17 normal full chest x-rays and 16 single lung images are shown to the participants in random order. Time of the diagnosis and the eye movements are also recorded using a remote head free eye tracker.

Results indicated that Expertise and CS play important roles in the diagnosis of pneumoconiosis. Novices and medical students are slow and inefficient whereas, residents and staff are quick and efficient. A key finding of our study is that the presence of CS information alone does not help improve diagnosis as much as learning how to use the information. This learning appears to be gained from focused

[1]In the context of viewing images, *Perception* is defined as the unified awareness of the content of a displayed image that is present while the stimulus is on [50]. *Cognition* is a group of mental processes that include attention, memory that understands, analyzes and makes decisions. Perception and Cognition are highly interrelated [82] and its difficult to distinguish factors affecting only the perception from those affecting only the cognition of observers (readers of chest x-rays). We didn't attempt to analyze the perceptual and cognitive factors separately, in the present study. But, we believe that most factors considered in the present study affect both perception and cognition of the observers.

training and years of experience. Hence, good training for radiologists and careful observation of each lung zone may improve the quality of diagnostic results. For residents, the eye scanning strategies play an important role in using the CS information present in chest radiographs; however, in staff radiologists, peripheral vision or higher-level cognitive processes seems to play role in using the CS information.

There is a reasonably good inter and intra observer fixation consistency suggesting the use of similar viewing strategies. Experience is helping the observers to develop new visual strategies based on the image content so that they can quickly and efficiently assess the disease level. First few fixations seem to be playing an important role in choosing the visual strategy, appropriate for the given image.

Both inter-rib and rib regions are given equal importance by the observers. Despite reading of chest x-rays being highly task dependent, bottom up saliency is shown to have played an important role in attracting the fixations of the observers. This role of bottom up saliency seems to be more in lower expertize groups compared to that of higher expertize groups. Both bottom up and top down influence of visual fixations seems to change with time. The relative role of top down and bottom up influences of visual attention is still not completely understood and it remains the part of future work.

Based on our experimental results, we have developed an extended saliency model by combining the bottom up saliency and the saliency of lung regions in a chest x-ray. This new saliency model performed significantly better than bottom-up saliency in predicting the gaze of the observers in our experiment. Even though, the model is a simple combination of bottom-up saliency maps and segmented lung masks, this demonstrates that even basic models using simple image features can predict the fixations of the observers to a good accuracy.

Experimental analysis suggested that the factors affecting the reading of chest x-rays of pneumoconiosis are complex and varied. A good understanding of these factors definitely helps in the development of better radiological screening of pneumoconiosis through improved training and also through the use of improved CAD tools. The presented work is an attempt to get insights into what these factors are and how they modify the behavior of the observers.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Medical imaging technologies form one of the most effective diagnostic tools in medicine. With the discoveries of seminal physical phenomena such as ultrasound, radioactivity, magnetic resonance etc., and the related advances in technologies enabling the development of modern medical instruments harnessing these physical phenomena, medical imaging has made a long stride forward during the past century [16]. Medical imaging can be used for probing into the structure, function and pathology of human beings and also other living organisms. We are able to picture the insides of living beings with such precision and details which were not even dreamt of a few decades earlier. Medical imaging tools are also being used for planning treatment and surgery as well as other imaging in biology.

Despite the advent of modern imaging technologies like PET (Positron Emission Tomography), MRI (Magnetic Resonance Imaging) and Nuclear medicine, Chest X-rays still are most widely used diagnostic tools in medicine, because of their low cost, ease of use and low x-ray dosage. Chest radiographs remain the most common and widely used tool for the diagnosis of lung diseases despite the impressive technical advances during the past four decades. Chest x-ray is still ubiquitous in clinical practice, and will likely remain so for quite some time [88].

The information provided by a medical image is itself not sufficient for diagnosis and treatment. This information has to be interpreted by humans in an accurate and timely manner, for proper diagnosis of the diseases and for treatment. Even though, the advances in machine learning and image processing techniques help us in developing good CAD (Computer Aided Diagnostic) tools, we still are not able to replace human observers with machines. So, radiologists still play a very important role in interpreting the medical images, which is the key for proper diagnosis.

Several factors come into picture, when observers read medical images. Not only the observer independent factors such as image quality and viewing settings, but also several observer dependent cognitive and perceptual factors play a very important and significant role in the accurate reading of medical images. The present work deals with the understanding of some perceptual and cognitive factors involved in the diagnostic assessment of a diffused lung disease called Pneumoconiosis (section 1.1), which is mainly diagnosed by reading chest x-rays. Specifically, we are interested in understanding the role of role of different factors such as expertize, symmetry in chest x-rays, image features on the visual perception and cognition of x-ray images by the observers.

In this introduction chapter, we will discuss some important aspects of Pneumoconiosis and its diagnosis. Then in section-1.2, we discuss role of perception research in medical imaging followed by the thesis objectives in section-1.3. Finally, we conclude this chapter by outlining the organization of this thesis in section-1.4.

1.1 Pneumoconiosis

Pneumoconiosis is the inflammation of the lungs. The principal cause of the Pneumoconiosis is prolonged work-place exposure for many years that causes patches of irritation to form in one or both lungs [61]. This results in the formation of scar tissue making lungs less flexible and porous. Not everyone exposed to the dusts actually become ill and it usually affects men over age 40 and usually takes at least 10 years of exposure and sometimes up to 25 years to show signs of the disease. The symptoms include shortness of breath, cough, restless sleep, chest discomfort and the nails and lips may appear pale or bluish due to poor oxygenation.

There is no specific treatment for pneumoconiosis. A general treatment can only relieve the symptoms of pneumoconiosis. Some treatment options include medication, removal of the patient from the workplace, providing dust control through added ventilation, or the use of personal protection devices like dust masks. Like in many other diseases, prevention and education are very important for controlling pneumoconiosis. Given the fact that most people getting affected by pneumoconiosis are workers, they lack sufficient education and safety measures. So, the effective way to prevent the progress of this disease is to get regular radiological checkups. Despite the increasing safety measures and better work place environments, pneumoconiosis is deemed to be the most common and serious occupational lung disease in the developing countries like China and India [89].

1.1.1 Diagnosis of Pneumoconiosis

Diagnosis of this disease is done through radiological observation of chest radiographs. The diagnosis of pneumoconiosis is a complex procedure and it requires a certain level of expertise [65, 75]. Some studies proved that there is significant inter reader and some intra reader variation in the diagnosis of pneumoconiosis [75, 44]. So, in general, there will be a hierarchy of readers for this disease. International labor Organization (ILO) has also introduced a standard classification scheme for the diagnosis of pneumoconiosis to facilitate the international comparisons of data, epidemiological investigations and research reports [21].

Disease Staging

ILO classification scheme [21] provides guidelines for classifying both parenchymal and pleural abnormalities and it involves classification of x-rays based on various parameters: quality of the image, profusion level, affected zones of the lungs, shape and size of opacities. For brevity, we discuss only the important aspects of the ILO classification required for the understanding of this thesis. The reader is advised to look at [21] for complete ILO classification scheme for Pneumoconiosis staging.

From clinical and occupational health viewpoint, the classification of radio opacities as to profusion category is of primary importance. Other information regarding opacities is of more epidemiological than clinical or compensative value. The profusion of radio opacities refers to the concentration of small opacities in affected zones of the lung. The category of profusion is based on the comparison with standard radiographs. The standard radiographs define four categories (0, 1, 2 and 3) with each category having 3 sub categories. Figure-1.1 shows x-ray segments of different profusion categories.

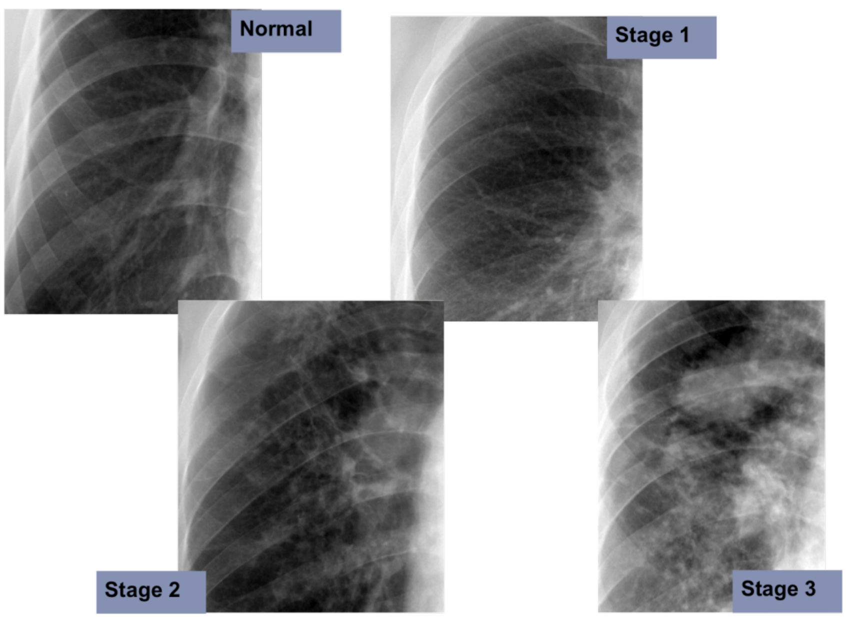


Figure 1.1 Sample X-ray segments showing different disease stages in Pneumoconiosis. (Source: Shanghai Pulmonary hospital, China)

From clinical and occupational health viewpoint, the classification of radio opacities as to profusion category is of primary importance. Other information regarding opacities is of more epidemiological than clinical or compensative value. The profusion of radio opacities refers to the concentration of small opacities in affected zones of the lung. The category of profusion is based on the comparison with standard radiographs. The standard radiographs define four categories (0, 1, 2 and 3) with each category having 3 sub categories. Figure 1.1 shows x-ray segments of different profusion categories.

Each lung is divided into three zones: upper, middle and lower by horizontal lines drawn at one third and two thirds of the vertical distance between the apex of the lung and the dome of the diaphragm

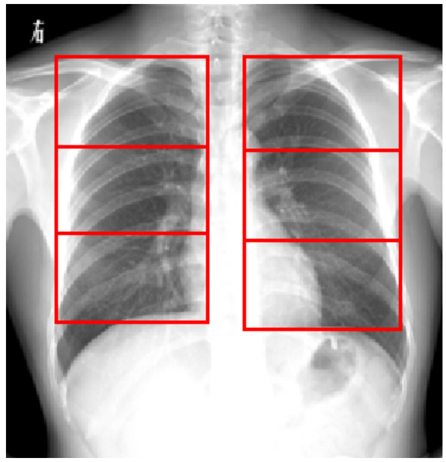


Figure 1.2 Division of lung fields into zones

(Figure 1.2). Then profusion for each zone will be noted. The overall profusion is determined by considering the profusion as a whole over the affected zones of the lungs. When there is marked difference in profusion in different zones of the lungs, then the zone/s showing the marked lesser degree of profusion is/are ignored for the purpose of classifying the overall profusion. Overall profusion level, size of opacities and number of affected zones are taken into consideration while staging the disease. Since, pneumoconiosis is a lung disease, more emphasis is given to given to lung regions and that too inter-rib regions while assessing the disease.

1.2 The importance of perception research in medical imaging

The role of medical imaging is to provide information and to have better diagnostic accuracy. *Perception* is defined as the unified awareness of the content of a displayed image that is present while the stimulus is on [50]. *Analysis* is determining the meaning of the perception in the context of the medical problem that initiated the acquisition of the image [47]. Even though, a lot depends on the type and presentation of medical images, perception and analysis of the presented image content forms the key to the proper diagnosis.

Until 1940s, it was taken for granted that what radiologists perceive and interpret is the faithful representation of the image content [50]. The role of perception and cognition in the process of medical image interpretation was not generally questioned [45] until a series of studies in late 1940s [27, 28, 20], which showed the presence of inter-observer and intra-observer variations of the radiologists interpretation of medical images, establishing the importance of the role of perception and cognition in the medical image interpretation.

It is well proven fact now that a radiologist's interpretation of medical images is highly subjective and is dependent not only on the acquisition and presentation of image content but also on several perceptual and cognitive factors. The factors affecting the radiologists image interpretation can be broadly classified into two kinds [60]:

1. *Image dependent factors*, which are related to the visual conspicuity of features relevant to the clinical problem; and

2. *Image independent factors*, which are primarily cognitive in nature and relate to what the observer knows about the visual information in front of him.

Although, development of new acquisition and viewing technologies helps in leveraging the diagnostic errors caused due to image dependent factors, perception research is very important to understand the image independent factors. At least, half of the errors made in clinical practice are perceptual in nature [46]. Thus understanding these perceptual and cognitive factors involved in the interpreting the medical images can help in improving the diagnostic performance of radiologists in many ways. Better understanding of these factors helps in the development of better image acquisition and viewing systems tailored to the perceptual needs of the observers, development of computer aided diagnostic (CAD) tools and also the development of better training regimens for the resident radiologists. As mentioned in [46], the goals of the perception research in medical imaging can be broadly outlined into following categories:

- Modeling the detection task
- Understanding the visual search
- Understanding the nature of expertize
- Developing perceptually based standards for image quality
- Developing computer aided perception tools
- Developing quantitative methods for describing natural images and for measuring human detection and recognition performance

The ultimate aim of all perception research is to improve the diagnostic performance by reducing the observer errors due to the perceptual and cognitive factors and in turn help in better patient care and treatment. The medical image perception research is still a relatively new and fast growing research area and there is still a lot of research work to be done in order to get a good understanding of the complex processes involved in the task of image analysis.

1.3 Objectives and Challenges

The present perception study is aimed in understanding the perceptual and cognitive factors effecting the assessment of pneumoconiosis through the reading posterior-anterior (PA) chest X-ray images [1]. This is done through a gaze tracking experimental study aimed in getting insights into the role of different factors on the diagnostic performance of the observers. Specifically, we are interested in answering the following research questions:

1. What is the role of *expertize* and *contralateral symmetry information* present in the chest x-rays on the diagnostic error, time and the eye movements of the observer?
2. Does the *distribution of eye fixations* change with observer error and observer assessment of pneumoconiosis?
3. What is the *inter observer and intra observer variability of eye fixations*?
4. What is the*role of anatomical features* in attracting the gaze of the observers?
5. What is the *role of bottom up image features* in attracting the gaze of the observers?
6. How do the *visual strategies* of the observers of different expertize levels change with time?

In addition to trying to answer the above research questions, the present study is aimed at developing a gaze predicting model which can predict the gaze of the observers to a reasonably good accuracy.

Some perceptual research work has already been done previously, on the diagnosis of Pneumoconiosis through reading digital chest x-rays. This related work is briefly explained in section-2.1. Some studies [75, 44, 65] showed the significant inter-reader and some intra-reader variability of the pneumoconiosis disease assessment. The present study is aimed in understanding the variability of eye fixations, but not the variability of disease assessment. Even though these studies [75, 44, 65] show the importance of expertize , to our knowledge, there are no studies on how the different levels of expertize affect the diagnosis of the disease. The present study is aimed in delineating how the different levels of expertize affect the assessment of pneumoconiosis. We didn't find any previous gaze tracking studies on the diagnosis of pneumoconiosis. The present work is an attempt in understanding the various factors (image features, expertize etc.) affecting the diagnosis of pneumoconiosis, through quantitative analysis of eye fixations.

The challenges in the present perception study are similar to the challenges of any experimental study. The experiment has to be designed and conducted in such a way that there would be no systematic errors and little random errors. The higher challenging task would be to have the following three *validities* [78] for our experiment:

[1]Most chest x-ray films are taken posterior anterior (PA). That is, the x rays shoot through from the back of the patient to the x-ray plate in front of the patient

- ***Construct Validity***: It is the extent to which a variable actually reflects the theoretical construct that we intend to measure
  - Are we really manipulating the independent variables such as contralateral symmetry and expertize that we intended to manipulate?
  - Are our measures clear representations of our dependent variables such as observer error?
- ***Internal Validity***: It is concerned with correctly concluding that an independent variable is, in fact, responsible for variation in the dependent variable
  - Controlling the effect of nuisance variables such as screen resolution, room conditions etc. on the dependent variables
- ***External Validity***: It is the extent to which the experimental results can be generalized to different contexts and individuals
  - Making sure that the experimental results in our experiment can be generalized to different radiologists and different clinical settings

We hope that we have taken all necessary measures, during experiment design, to maintain the above three validities as much as possible in our experiment. But, no experiment is perfect. For example, it is difficult to quantitatively measure some variables such as *observer error, inter and intra observer variability* etc. since there is no standard method for measuring these behavioral aspects. But, we believe that we have made sufficient justifications for our methodologies and measures used in our experiment.

Another important challenge in our study entailed the development of a new saliency model that can predict the gaze of the observers while assessing Pneumoconiosis. Even though researchers have made significant progress in the understanding of human visual system, we still dont know exactly how the various top down and bottom up influences are affecting the gaze of the observers [15]. For instance, we do not know what exactly constitutes a top down influence of visual attention and we do not know how the bottom up and top down influences interacts with each other. We made some assumptions, whenever necessary, to develop our saliency model which can predict the gaze of the observers to a reasonably good accuracy. Sufficient justifications are also made for the assumptions made.

1.4 Contributions

The following are the main contribution of the thesis:

1. Insights about the
   (a) role of contralateral symmetry and expertize on profusion ratings of observers of varying expertize levels.

   (b) visual search strategies of observers with varying levels of experience.
   (c) importance of anatomical structures in chest X-ray image perception.

2. Verification of the predictability of fixations by bottom up saliency models.

3. An extended bottom up saliency model for predicting eye fixations.

1.5 Thesis organization

In this chapter-1, we introduce the reader to Pneumoconiosis and its diagnosis; discuss the importance of perception research and; also discuss the basic objectives and challenges in this thesis work. In chapter-2, we will see the previous work along with some other research work required for the better understanding of this thesis. The present work being a perception research on Pneumoconiosis diagnosis, we discuss some existing perception research on Pneumoconiosis and the chest x-rays in general. We provide some basics of visual attention and some models of visual attention, as the present work is mainly an eye tracking study. In chapter-3, we discuss the important details regarding the experimental design and the experimental procedure, which is basically an eye tracking experiment on the observers ranging from novices to staff radiologists.

Even though, there are some studies showing the importance of expertize on contralateral symmetry, its role is completely not understood. In chapter-4, we analyze the role of expertize and contralateral symmetric information present in chest x-rays on the diagnostic performance of the participants of the experiment and also on their eye movements. In chapter-5, eye movements of the observers are analyzed in further detail to study the inter and intra-observer fixation consistency along with the role of bottom-up and top-down image features in guiding the fixations of the observers. This chapter also analyses how the visual strategies of observers of different expertize groups, change with time. In chapter-6, based on the experimental results, we introduce a new gaze predicting model by modifying the bottom-up saliency maps with segmented lung masks. This new gaze predicting model performs, with a reasonably good accuracy in predicting the eye fixations of the observers. Finally, we end the thesis with conclusions and some future work directions in chapter-7.

Chapter 2

Background

Reading the seemingly simple PA chest x-rays is a quite complex task and understanding the role of several factors in reading chest x-rays is even more complex. This chapter discusses the basic research work on Visual attention and perceptual research on reading digital chest x-rays, which are related to the present work. In section 2.1, we discuss the role of some factors affecting the assessment of pneumoconiosis that has been studied in the literature. In sections 2.2 and 2.3, we discuss visual attention, the role of bottom up and top down influences of visual attention; and various models of visual attention. In section 2.4, we discuss some important perception research work on chest x-rays, unrelated to pneumoconiosis, but is important for the better understanding of this thesis work.

2.1 Some factors affecting the diagnosis of Pneumoconiosis

Apart from reading digital chest x-ray images, patient working history plays an important role in assessment of the disease level. But, chest radiographs retain their paramount position in the diagnosis, investigation and management of the disease. Several factors have been shown to affect the diagnosis of pneumoconiosis, in the literature.

X-ray image quality plays a serious and significant role [75, 72]. There is a marked tendency to award higher readings to the under-penetrated or soft films, while the opposite was true of the over-penetrated or hard films. The experienced readers are less influenced and are more able to adjust for unsatisfactory film quality.

Diagnosis of pneumoconiosis is very subjective. Several cognitive factors play a very important role in the diagnosis. It is evident from the fact that there is substantial inter-reader variation and some intra-reader variation in the diagnostic assessment of pneumoconiosis [75, 44, 65]. The reader consistency from one time period to another is felt to be within acceptable limits, provided the reader is experienced. Even though these studies show the importance of expertise, there are not enough studies on *how* the different levels of expertise affect the diagnosis of the disease. The perceptual research work done on the diagnosis of pneumoconiosis is quite less and thus our present understanding of the various factors influencing the pneumoconiosis assessment is very limited. Since, understanding the role of contralateral symmetry (CS) present in chest x-rays on the diagnosis of pneumoconiosis is of interest, we discuss the importance of CS on reading chest x-rays.

2.1.1 Contralateral Symmetry and its role in diagnosis of Pneumoconiosis

Symmetry is present in many of the objects we encounter in our daily lives. The detection of symmetry is one of the characteristics of human visual perception. This may be due to the abundant examples of symmetry in the structure and development of the human and animal forms. For example, symmetry plays a very important role in visual processes such as face perception [18] and attractiveness [30]. Much of the gross anatomy of the human body also exhibits contra-lateral symmetry. Contra-lateral symmetry seems to be playing a major role in perception of things around us.

This might be also true with chest X-rays. PA Chest X-rays also exhibit good amount of contralateral symmetry (CS). Strictly speaking, lungs are naturally positioned in pseudo-contra lateral symmetry, but the near perfect contra lateral symmetrical occurrence of abnormal lung tissue in both lungs is highly improbable. There are good chances that radiologists use this contra-lateral symmetric information present in chest X-rays while reading them. In fact, radiological trainees were explicitly told to compare the left and right lungs while reading a chest X-ray. Prof. Kakarla Subba Rao, an eminent radiologist and former director of Nizam institute of medical sciences in Hyderabad, India, opines that the best way to read any chest X-ray, of any disease, is to compare the left and right lung zones [81].

Some observer studies [87] have shown the potential usefulness of *Contralateral Subtraction* [57, 95, 56] technique. Here, one side of the lung is subtracted from the other side after flipping and warping, so that the abnormalities stand our more clearly, either for the radiologist to see or for the computer to detect. Figure 2.1(a) shows the overall scheme of this contralateral subtraction technique. We are not discussing the details of the steps involved in this technique, here. Figure 2.1(b) and (c) shows a sample chest x-ray and the corresponding contralateral subtraction image. When the residual image is shown along with the original image, the detection accuracy of lung nodules was found to improve [87].

Despite the existence of the observer studies on the usefulness of contralateral subtraction technique, to our knowledge, there are no empirical studies, which study the role of contralateral symmetry in diagnosing lung diseases. Moreover, unlike lung nodules (which are localized), pneumoconiosis is a diffused lung disease. The precise role of contralateral symmetry in the detection and classification of pneumoconiosis is still not known and this is one of the main aims of the present study. Since eye tracking and gaze analysis are central to the present study, we discuss some basics of visual attention and some models of visual attention, next

2.2 Visual Attention

Visual search is one of the basic activities of human beings. We use visual search very frequently in our daily activities like searching for a book on the table etc. It is also one of the important activities that

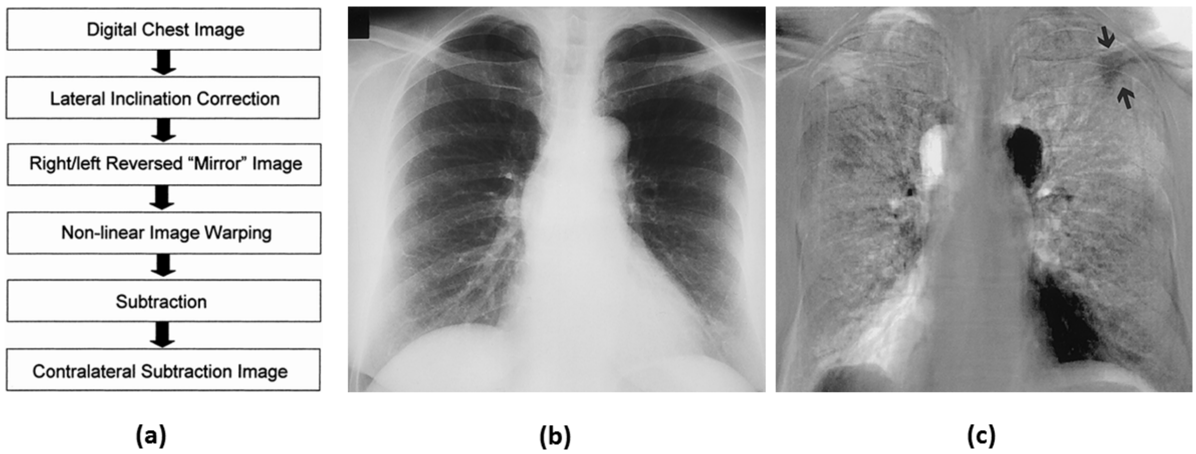


Figure 2.1 (a) Overall scheme of a contralateral subtraction technique for posterioanterior chest images. (b) A sample chest x-ray image and (b) its corresponding contralateral subtraction image. (Source: [87])

radiologists employ while searching for pathologies in medical images. This visual search activity is mediated by a perceptual and cognitive process called *visual attention*. Visual attention is the process of selectively attending to an area of visual field while ignoring the surrounding visual areas. Our human visual system, despite being sophisticated and highly developed, is limited in its resources and we cannot process the entire visual field in one instance. We have to actively attend to different areas of the visual field in order to successfully perform a given search task.

When searching an image, the visual system is involved in mainly two activities: examining the visual input for the target and scanning the eyes over the image. These activities are necessary because of the non-uniform sensitivity of sensory layout in the retina to stimuli. The human retina, a part in our eyes where light signals are converted into electro-chemical signals, consists of millions of photoreceptors called *rods* and *cones*. The distribution of these photoreceptors is not uniform across the retina and the density is more at the center part of retina called *fovea*. Thus the center of the visual field has the greatest resolving power and there is a gradual decrease in the visibility of objects in the periphery. This dominance of central vision is partly because of the structure of the retinal sensory array and partly because of neural connections to the visual cortex that favor the foveal cones.

The act of *selective visual attention* is divided into two processes namely *covert* and *overt attention* [93]. Over attention refers to the act of directing sense organs towards the stimulus source. That is, moving our eyes and heads towards the objects of interest. Covert attention is the cognitive act of mentally focusing on an object in the periphery of human visual field. Unless, we consciously concentrate on objects in peripheral visual field, it is the *overt attention*, which plays major role in selective visual attention. Covert attention is thought to be a neural process that enhances the signal from a particular part of the visual field. See [90] for a good discussion and examples related to visual attention.

From the above, it is clear that eye movements play a very important role in human visual attention. The following are the 3 neural regions implicated in eye movement programming and their functions [24]:

- *Posterior Parietal Complex*: Disengages attention.
- *Superior Colliculus*: Relocates attention.
- *Pulvinar*: Engages or enhances attention.

Although conscious control can be exercised over eye movements, this eye movement control is usually unconscious. In general, where radiologists attend to in medical images differs from what they think they have attended to. Thus, eye tracking research is very important to understand the various processes involved in our visual attention. Visual attention is a vast area of research and, here, we discuss only some basic and important aspects that are relevant to the present work.

2.2.1 Top down and Bottom up influences

What determines what we attend to? The perceptual and cognitive factors influencing the visual attention can be broadly classified as *top-down* and *bottom-up influences*.

The Bottom-up influences are factors that are dependent on the features of the visual stimulus and are independent of the observer. Irrespective of the knowledge of the observer and the task at hand, there are certain aspects of the visual stimulus that have the tendency to attract the gaze of the observers. For example, bright warning signals on road, red rose in an otherwise complete green grass etc. Basically, image features such as color, contrast and orientation have tendency to attract the gaze of the observers. The attention due to these bottom-up influences is also called *stimulus driven* or *exogenous attention*.

Top down influences are the image independent factors such as the given task or goal and knowledge of the observer that influence the process of visual attention. Yarbus [94] did some first eye tracking studies showing the importance of task on visual attention. Figure 2.2 shows a classical example of how the task given to the observer affects his/her eye movements. Notable are the differences of eye movements corresponding to tasks: 'Free examination', 'Estimate material circumstances of the family' and 'Give the ages of the people'. During 'free examination' task, the observers were scanning over the entire scene. Whereas, when observers were asked to give the ages of the people, they were mainly looking at the people in the scene. Top-down influences include all kinds of factors concerning the mental state of observer and knowledge of the outer world. This includes aspects like prior knowledge of the scene or of the objects that might occur in the environment and emotions, desires, intentions, motivations.

These influences of attention are image independent and are called top-down influences. Attention due to these influences is also called *goal-driven, endogenous attention* or *executive attention*. Even though the relative role of the top down and bottom up influences in modulating the visual attention is still not completely known, both play a very important role in human visual attention [43].

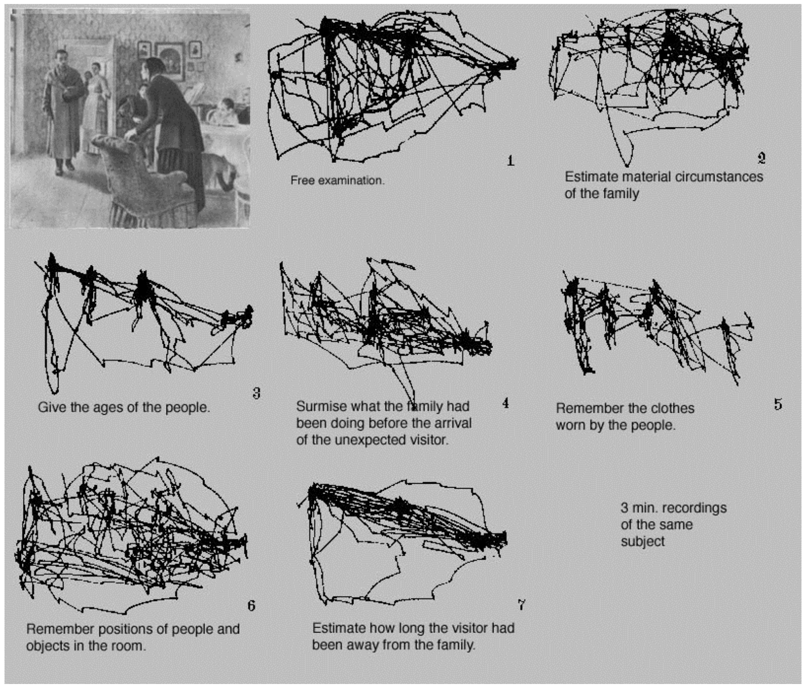


Figure 2.2 Eye movements of an observer over a picture (top-left) while performing different tasks. (Source: Yarbus, 1967 [94])

2.3 Models of Visual Attention

Several models of visual attention have been proposed in the literature. Their main objective is to simulate the behavioral data and to better understand human perception. Here, we only review only a few models which are influential and cover the majority of approaches and ideas found in psychophysical and computational modeling of visual attention.

2.3.1 Psychophysical Models of visual attention

Psychophysical models of visual attention are theoretical models explaining the process of visual attention. They try to explain the psychophysical dynamics of visual attention by systematic study of human eye movement behavior. Several psychophysical models of visual attention has been proposed in the literature. We discuss only two important models here: ‘Feature Integration Theory’ and ‘Guided Search Model’.

Feature Integration Theory (FIT) [85]

This is one of the most influential theories on visual attention which mainly posits that visual attention is responsible for binding different features into consciously experienced whole. According to this theory, the perception of an object has mainly two stages:

1. *Pre-attentive stage*: During this stage the object is analyzed in terms of its different features such as color, shape, orientation etc., which are processed in different areas of the brain.

2. *Focused attention stage*: This stage integrates different features in order to perceive the object as a whole, or recognize it if enough information is available.

Thus, according to FIT, attention is responsible for binding various features of an object to perceive or recognize the whole object. Information from different feature maps are collected in a master map (also called *saliency map*). For example, when searching for a green circle among different geometrical shapes of different colors, neither the color feature 'green' nor the shape feature 'circle' is sufficient to located the search target. Several physiological and anatomical studies support the hypothesis that the different visual features of an object are encoded in different areas of our brain. This theory provides a plausible explanation of how these different features are integrated to perceive the object as a whole. A number of computational models of visual attention are also proposed based on this theory.

Guided search Model (GS) [91]

Wolfe's guided search model of visual attention is proposed as an answer to some of the criticisms on the early works of FIT. This model is proposed along the lines of FIT with multiple feature maps computed along parallel channels and are then integrated into a single map known as *activation map*. This model is more detailed in several aspects which make it more suitable for computer implementation. The basic difference between FIT and GS is that the former considers different feature maps for different features (blue, green etc.) in the same feature dimension (color, intensity etc.), whereas the GS considers only one feature map for each feature dimension. Another major difference is that the GS also considers the top-down influence in selecting and integrating the feature maps. Wolfe's popular Guided Search Model offers a more up to date theory of visual search compared to FIT.

Several other psychophysical models of visual attention such as Dynamic routing circuits [11], Selective attention model (SLAM) [74], Search via recursive rejection (SERR) [39], and Selective attention for identification Model (SAIM) [33] etc. have been proposed. [34] provides a good review of these various psychophysical models of visual attention.

2.3.2 Computational Models of visual attention

Compuatational models of visual attention provide the computational details of the process of visual attention so that it would be possible to implement them on computers. Many popular computational visual attention systems are biologically motivated. The output of any computational model of visual attention is a *saliency map*. Saliency map is a topographically arranged map that represents visual saliency of a corresponding visual scene. A saliency map can be represented as a grayscale image in which each pixel value represents the likelihood of the corresponding visual region in attracting the attention of the observers. All computational models of visual attention take a representation of visual field, such as an image; perform some processing and then output a saliency map corresponding to the input visual field. Different models differ mainly in how this processing is performed. The processing can be broken up into two broad components [80]: *feature extraction* and *gaze computation*.

Many computational models focus on extracting mainly three features [25]: intensity, orientation and color. These features are used mainly because they are proposed in the underlying psychophysical models such as FIT and GS. They are also relatively easy to compute. The Itti and Koch model [40], one of the most popular computational models of visual attention, uses these three features. Many other high level features such as curvature [63], optic flow [86], symmetry [14] etc. are used in the literature for computing saliency map. Some recent models also successfully use spectral based features [35, 36] for saliency computation. Different models proposed in the literature have slightly different objectives and contexts in which they are used. Features are selected according to these contexts and objectives. For example, attention models developed in the context of object detection [35] would use different features than those developed in the context of free viewing [40, 32].

The gaze computation step involves the computation of gaze location from the abstractions computed in the feature extraction step. This gaze computation step can be broadly divided into two steps [32]: activation and normalization/ combination. The activation step forms activation maps using the feature vectors and normalization/combination step normalizes the activation maps followed by a combination of the maps into a single map. These steps are discussed in more detail when we discuss the Itti-Koch model below. Even though feature extraction is the main step in saliency computation, [32] showed that activation and normalization steps also play an equally important role. To understand the general steps in a computational attention system, we discuss the Itti-Koch [40] model of visual attention in more detail.

Figure 2.3 shows the basic structure of the Itti-Koch model [40]. This model is a derivative of Koch-Ullman model [42] and is one of the most popular attention systems. From an input image, three features (color, intensity and orientation) are computed at 8 scales to build an *image pyramid*. The *center-surround differences* and *normalization* step finds the conspicuous regions related to different features and forms activation maps for each feature dimension. These activation maps are combined in a *linear combination* step to get a central saliency map. A *winner take all (WTA) network* determines the most salient region in this map which yields the focus of attention. The details regarding these different steps are not discussed here. Several other models such as GBVS saliency model [32], used this basic architecture of Itti-Koch model with some modifications to different steps.

Most of the leading computational attention systems, in the literature, are designed to detect the bottom-up salient regions. This is because of the difficulty in conceptualizing various top-down influences such as prior-knowledge, emotions etc. The output *saliency maps* in Itti-Koch model and many popular saliency models are also called *bottom-up saliency maps* as they are computed solely based on image features.

The Guided search model [91] hypothesizes that this bottom up processes, used in many computational models, can be biased for features and locations thus accounting for top-down influences of visual

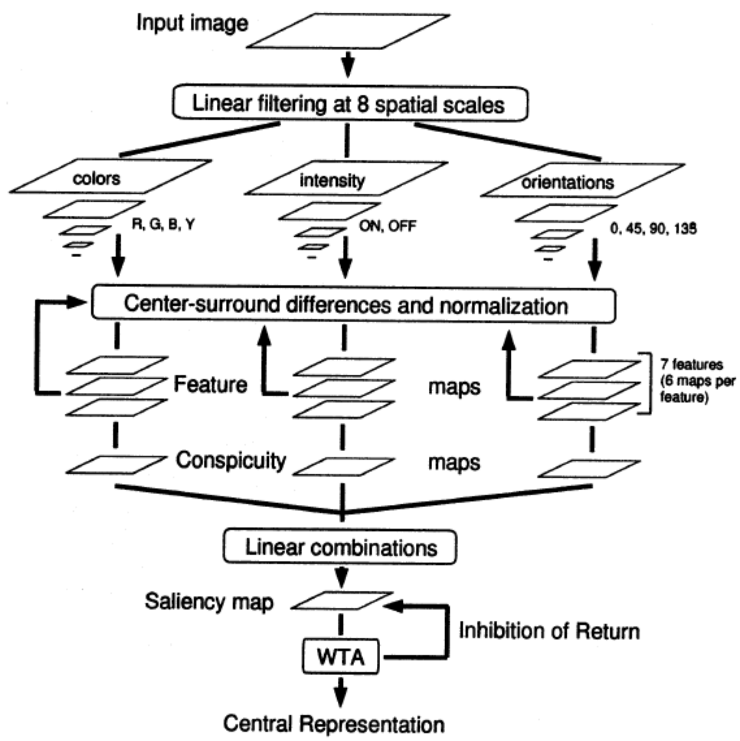


Figure 2.3 Flow chart showing different steps in Itti-Koch computational saliency model (source: [40])

attention. Following these lines, some computational systems have been developed with top-down information (such as prior-knowledge of scene) influencing the bottom-up processing at different stages: some systems influence *feature extraction* step [86, 68], some influence *gaze computation* step [64]. Some systems compute the bottom-up saliency and then just investigate the computed salient regions for target similarity [67], where target refers to the features of the object to be identified in a visual scene.

Some recent systems [73, 76], instead of tuning the bottom-up processing, attempt to model top-down influences by computing a separate *task relevance map*, and then combine the bottom-up saliency map with the task relevance map. Task relevance map might be computed by combining information about desired features, cued spatial locations, scene gist and context etc. Some recent human neuroimaging data support the existence of such task relevance maps in the intraparietal sulcus (IPS) [25]. Peters & Itti [73], computed task relevance map by learning the associations between some low level image features and human eye movements of the observers while engaged in a driving task. This task relevance Top-down (TD) saliency map is then combined with bottom-up (BU) saliency map, computed using Itti-Koch saliency model [40], giving rise to a final BU*TD priority map that guides attention. Figure 2.4 is the schematic illustration of this model [73]. In the *training phase*, the low level feature vectors (pyramid features, Fourier features etc.) are extracted from the training set of clip images and are then passed to a machine learning algorithm to learn a mapping between feature vectors and eye positions. Then, in the *testing phase*, frames from the test clip are passed in parallel to a bottom-up saliency model as well as to the top-down feature extractor that is used to generate a top-down eye position prediction map (*task relevance map*). Finally, these BU and TD prediction maps are combined, through point-wise multiplication, to obtain final saliency map that predicts the eye movements of the observers.

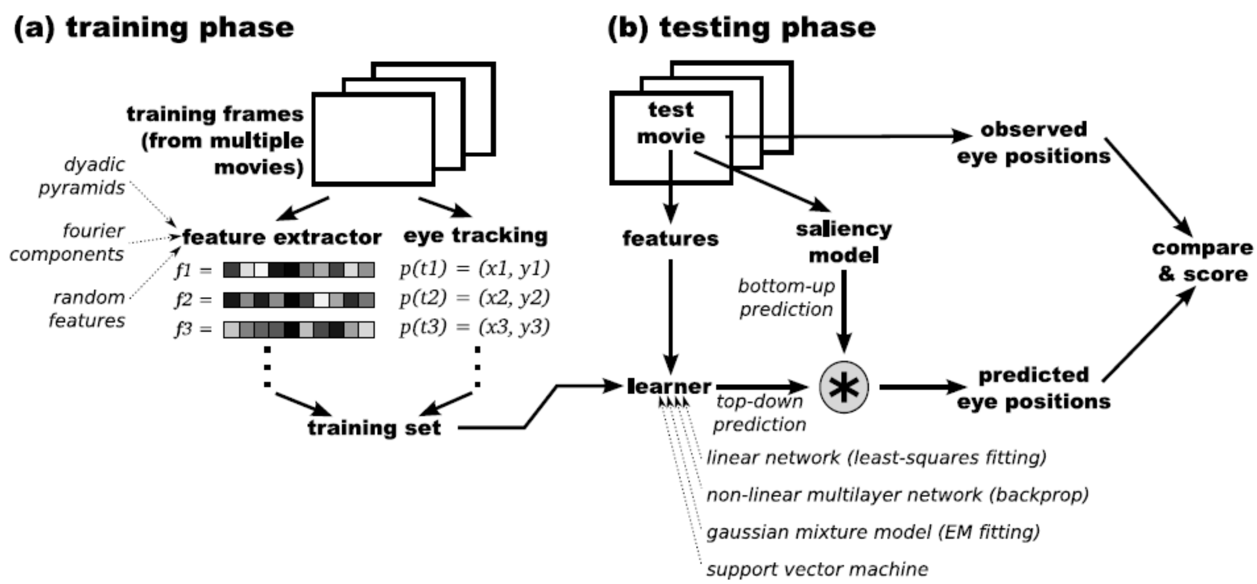


Figure 2.4 Illustration showing the steps in Peters and Itti's computational model (source: [73])

Even though, some recent advances have been made in incorporating the top down influences in computational models, currently, there is no complete and robust system which analyzes the role of top-down attention for different contexts and covering different aspects of top-down influences. The relative roles of top-down and bottom-up influences and how they interact with each other in guiding the human eye movements remain unclear.

In section-2.1, we disscused some existing perceptual research on the diagnosis of pneumoconiosis. Next, we discuss some perception research on reading digital chest x-ray images.

2.4 Perception research on Chest x-rays

Most of the existing perception research on chest x-rays is done in the context of radiologists searching for localized lung diseases such as lung tumors. Nevertheless, being perceptual experiments on chest x-rays, this existing research on localized lung diseases played an important role in the design of the present perception study. So, it would be good to discuss briefly some important and relevant points, in the literature related to perception research on chest x-rays, even though they are not related completely to Pneumoconiosis, for better understanding of the present work.

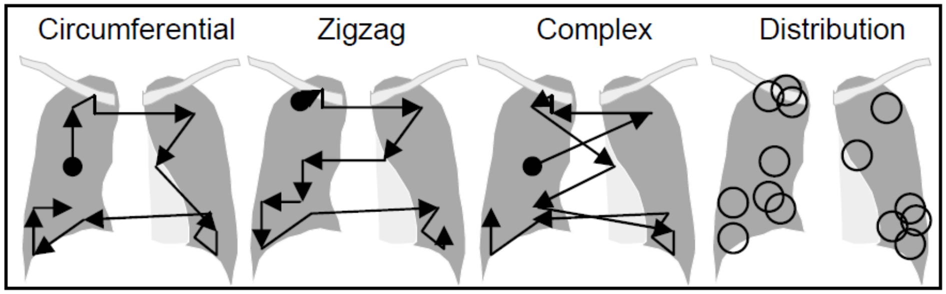


Figure 2.5 Different scan paths with same fixation distribution. (source: [49])

2.4.1 Eye movement research

Visual search plays a very important role in reading medical images [48]. Studying the locations where the observers are attending to in chest X-rays gives good insights into the perceptual and cognitive mechanisms involved in reading x-rays. Such studies are generally done by an *Eye Tracking* process. Before discussing the important eye movement research related to chest x-rays, we briefly discuss the eye tracking process below as this forms the core of present study.

Eye Tracking: Eye tracking is a process of measuring the point of gaze over a given visual scene. This is done by using *Eye Trackers*, which are devices for measuring eye positions and eye movements. There are number of different types of eye trackers. The most popular and recent eye trackers use video images from which the eye position is extracted and recorded. A camera focuses on one or both eyes and records their movements as the viewer looks at some kind of stimulus. Most modern eye trackers use infrared light to create corneal reflections [31]. Then the gaze direction is computed using the vector between the pupil center and corneal reflections. A simple calibration procedure is usually needed before using the eye tracker.

Some important results related to the eye movements of the observers while searching for lung nodules in chest x-rays are reported in the literature. We summarize them next.

- ***Less sampling by fovea***: It has been observed that large areas in chest x-rays are not sampled by the fovea of the radiologists [48, 38]. This does not necessarily mean that search is inefficient but it does illustrate the importance of parafoveal vision in the image analysis.

- ***Non-random fixation patterns and visual search strategies***: Radiologists move eyes in a pattern that is neither random nor the same as that of a layman. Consistent initial search strategies are found in trained viewers if the search task is clearly specified [55]. There is a tendency to fixate upon edges and to exclude broad uniform areas [55, 84].

- ***Evolution of fixation patterns with expertize***: There is a definite evolution of fixation pattern from that of an untrained person to that of a radiologist, which occurs during the medical school and changes very little during residency training [51]. The development of search strategy and of an ultimate fixation pattern depends more upon knowledge of radiographic anatomy, pathology, and clinical medicine than upon formal radiologic training as given in a residency program.

- ***Similar distribution of eye fixations by different radiologists***: Visual scanpaths on chest x-rays can be broadly classified as circumferential, zigzag and complex [49]. These are shown in figure 2.5. More than half of the scans of radiologists are too complex to classify. But, it has been found that despite following different scan paths, different radiologists tend to fixate over same locations [48, 49], as shown in the last image in figure 2.5. It is not clear if the sequence in which the visual information is collected by the radiologists is important for diagnostic performance.

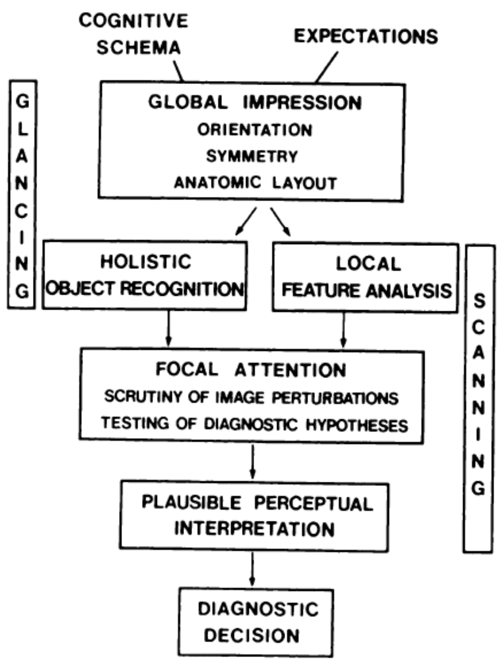


Figure 2.6 Global-focal detection model of visual search. (source: [70])

The global-focal detection model of visual search [70, 69]

Nodine and Kundel [70, 69] have developed a model of visual search and detection that has three main components: overall pattern recognition (global impression), focal attention to image detail; and decision making. Figure 2.6 shows a diagrammatic representation of this model. According to this model, visual search begins with a global response involving the entire retina, in which the context is established and gross deviations from normal are detected. This response initiates a series of checking fixations, using the fovea to resolve ambiguity and fill in detail. The total search strategy consists of an ordered sequence of interspersed global and checking fixations. We briefly discuss these three components.

1. *Overall pattern recognition*: The first glance at the image produces a global impression which provides the perceptual system with the information needed to carry out the diagnostic task. This global analysis identifies perturbations, which are novel and unexpected features, and sets the stage for the detailed focal analysis of the image. Most of the obvious abnormalities can be detected through global impression with careful analysis of different regions of the image.

2. *Scanning*: Following the global analysis, radiologists scan over the different image areas, scrutinizing different parts of the image with central foveal field. This careful analysis of different image regions is essential for detecting subtle abnormalities.

3. *Decision*: After carefully scanning various portions of the image, the radiologist would arrive at a plausible diagnostic interpretation of the image.

Even though the relative role of these three components in reading chest x-rays is completely not understood, several studies showed that all the three components play a very important role in detection of abnormalities in chest x-rays.

2.4.2 On visual dwell time

The analysis of time of diagnosis can also provide very useful insights into the cognitive processes involved in reading chest x-rays. Two types of analysis can be done regarding the time: 1. Analysis of time required to find a target and; 2. Analysis of time spent fixating the target (dwell time). *Dwell time* is the total eye fixation time within a zone of an image. Zones are usually circular or elliptical regions defined around the lesions.

Some studies [52, 71] showed that flash viewing (0.2 seconds) is sufficient for detection of large or obvious abnormalities in chest x-rays. Thus, even within the short viewing time, too small for eye movements, radiologists are able to recognize the large abnormalities. This shows the importance of peripheral vision in reading chest x-rays. On the other hand, a substantial portion of subtle lung lesions are missed even with unlimited viewing time [71]. The detectability of lesions decreases considerably as viewing time becomes less than 4 seconds. These studies show that the chest x-rays should not be speed read.

An experimental study on chest x-rays [19] showed that there are two components of perception, working side-by-side while reading chest x-rays: Rapid and slow components of perception. During *rapid component* of perception, obvious abnormalities are detected instantly by comparing the radiograph with previous learned concept of normal. The detection time of these abnormalities is so small that systematic search seems unlikely. The study [19] hypothesized that the more the experience, the greater the number of abnormalities detected via this method. The *slow component* of perception is dependent on search, which is essential for detecting subtle abnormalities. The longer the search, the greater would be the number of observations.

2.4.3 Observer Error

The errors observers make in detecting and interpreting targets in medical images can be classified into 3 categories [70]: sampling errors, recognition errors and decision-making errors.

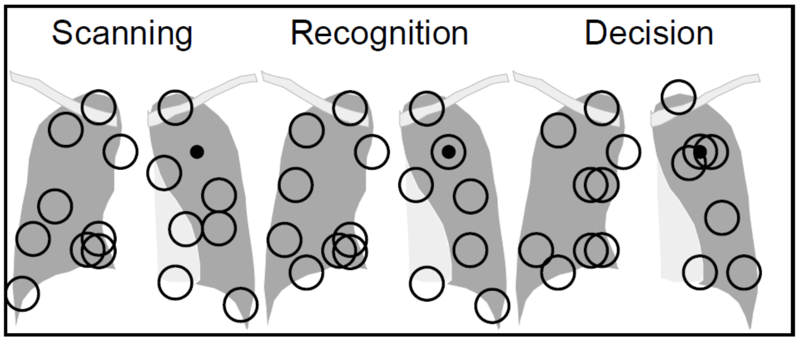


Figure 2.7 Three categories of error as determined by the analysis of scanning over the image. The black circles indicate human fixations and the black dot indicates a target. (source: [49])

Sampling errors: These are the errors made when the observer failed to fixate the lung nodule. These are also called scanning errors. If a target was not hit (not located within 2.8 degrees of center of fixation), a scanning error was scored. See Figure 2.7 (left), where the black dot represents a target and black circles represent observer fixations. As shown in this figure, since the target is not fixated by the observer, the error is considered a sampling error. In general, around 30% of the observer errors are due to sampling error [53]. Mechanical scanning aids might help in reducing this error.

Recognition errors: Many targets are not reported even when they are fixated upon. If the target was hit for one fixation and the gaze did not return, a recognition error was scored. See figure 2.7 (middle), which shows the target being fixated once. In general, 25% of observer errors fall into this category [53]. A well specified instruction might help in reducing these errors.

Decision errors: Even after carefully scrutinizing a target, the observer may not report it thinking it as a variant of normal tissue. Such errors are called Decision errors. If the gaze was prolonged or if the gaze returned to the target one or more times, a decision error was scored. See Figure 2.7 (right). In general, 45% of the errors are due to decision making [53]. Thus, the large amount of errors in this category shows the importance of decision making i.e. deciding whether an abnormal looking region is in fact abnormal or not.

2.4.4 The importance of Expertize

Expertize is considered to be the most important factor, in reading medical images. Several perception studies on chest x-rays also shows that this is true. We have already discussed some important results related to the effects of expertize on scan patterns and time of diagnosis. Several studies [51, 49, 71, 53] showed that experts seem to make more *efficient use of information from the peripheral retina* and fixate abnormal locations in the image more quickly and efficiently than non-experts.

In general, experienced staff radiologists are quick and efficient whereas inexperienced and resident radiologists are slow and inefficient [19].

Based on experimental results of a study on expertize [66], it was hypothesized that expert radiologists appear to view x-ray images, in the same way that we process faces i.e. by quickly detecting and processing the features that distinguish one stimulus from another. With experience, the experienced radiologists seem to develop the ability to detect the abnormalities and at the same time, develop ability to ignore the variations in normal features.

Regarding the effect of expertize on visual search strategies, an experimental study [51] on fixation patterns showed the evolution of visual search strategies from untrained observers and medical students to residents and staff radiologists. Kundel and Paul [51] found some characteristic differences in search strategies of observers belonging to different expertize groups. They found that, in general, radiologists fixations have broad coverage of the x-ray film whereas, untrained observers fixations were more clumped towards central portions of the image. The search strategies of the resident and the medical student groups are found to be intermediate between those of trained radiologists and untrained subjects.

2.5 Concluding Remarks

In this chapter, we have discussed some previous work, which is related to present research work. We mainly discussed the basics of visual attention, its models and roles of top-down and bottom-up influences. We also discussed perception research on Pneumoconiosis diagnosis along with some important results related to perception research on chest x-rays. Very little perception research work is done on diffused lung diseases like that of pneumoconiosis. Apart from the effects of expertize and x-ray image quality, we still do not know the role of other perceptual or cognitive factors on the diagnosis of pneumoconiosis. Although significant research is done on chest x-rays of localized lung diseases, we do not know which of these concepts concerning the localized lung diseases would be valid for diffused lung diseases like pneumoconiosis. The present study is aimed in getting some insights into the perceptual and cognitive factors affecting the diagnosis of Pneumoconiosis.

Chapter 3

Methodology

Behavioral experiments are one of the best ways of getting insights into the perceptual and cognitive processes of the radiologists. In the last chapter, we have seen that very less perception research has been done on the diagnosis of pneumoconiosis. Given the diffused nature of this disease, we believe that the perceptual and cognitive factors involved in the diagnosis of pneumoconiosis would be different from the factors involved in the diagnosis of a localized lung disease. The present study is aimed in understanding the various factors affecting the diagnosis of pneumoconiosis.

As discussed in the previous chapters, diagnosis of pneumoconiosis is a challenging task and several perceptual and cognitive factors seems to be affecting the reading of chest x-rays of pneumoconiosis. Specifically, the present study is aimed in understanding the role of expertize and contralateral symmetric (CS) information present in chest x-rays, on the diagnosis and the eye movements of the observers. We are also interested in understanding the image features that attract the attention of the observers. We have designed an eye tracking study to understand the role of various factors on diagnosis of pneumoconiosis and thus meet our objectives explained in section-1.3. Eye tracking is the best way to extract information on the reading pattern and hence is our chosen method of study. The present chapter gives details regarding the eye tracking experimental design.

3.1 Experimental Details

Our experiments were conducted in a room, dedicated to eye tracking experiments. Written informed consent, for the study and eye movements recordings, was taken from all the participants. The chest X-rays used for the experiment were collected between 2006 and 2008, from the Shanghai Pulmonary Hospital, China; and provided by GE Global Research, Bangalore, India.

3.1.1 Experimental Images

In order to understand the role of contralateral symmetric information present in x-rays on the disease assessment, we showed both single lung and double lung images to the observers. The test image

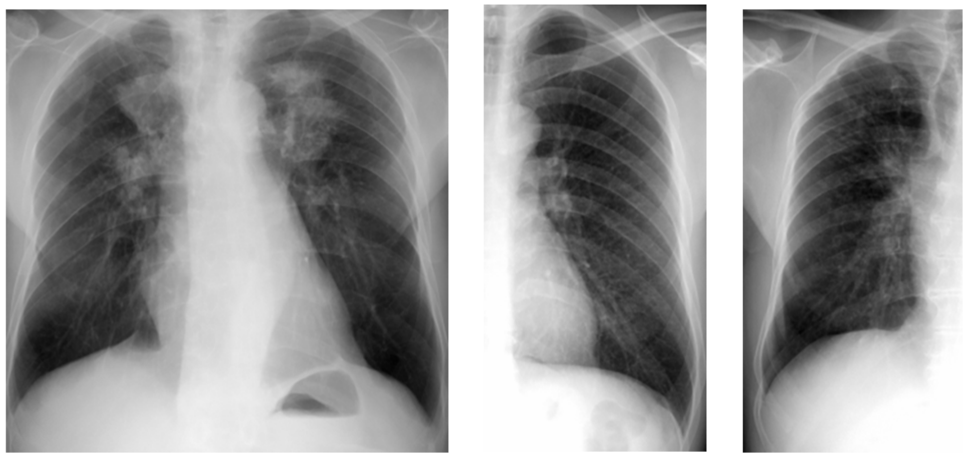


Figure 3.1 Sample double (left) and single lung (right) images used in our experiment

set consisted of 33 good quality PA digital chest X-rays, of which 17 were normal, full, double-lung images and 16 were single-lung images. All the images were cropped to remove the unnecessary surrounding background regions. Final cropped images, used for our experiment, have a spatial resolution of 1024x1024 for double lung images and 512x1024 for single lung images. All images have gray-level resolution of 4096 (12 bits). Figure 3.1 shows the sample single and double lung images used for our experiment.

The x-ray images were of laborers working in an industrial environment with either silica or metal dust. The workers had spent anywhere between 2 years and 31 years (*Mdn* = 12) in this environment. The experienced radiologists at Shanghai Pulmonary Hospital, China provided ground truth for these images. The participants in the experiment were blind to the occupational and other clinical related information of the patients. Test images included images of disease stages: 1, 2 and 3. Disease stage of '1' corresponds to low level of abnormality and disease stage of '3' corresponds to high level of abnormality. Refer to Section 1.1 explaining the disease staging of Pneumoconiosis. Table 3.1 shows the distribution of test images in different disease stages.

Table 3.1 Images of different disease stages used in our experiment

Disease Stage	***Double lung Images***	***Single Lung Images***
Stage 1	3	2
Stage 2	4	6
Stage 3	10	8
	Total: 17	Total: 16

3.1.2 Participant Details

In this thesis, we mainly use the word 'observers' in referring to the human participants in our experiment. Since one of the purposes of our experiment is to study the role of expertize, we ran the experiment on observers varying from novices and medical students to expert radiologists. These observers were in 4 categories: staff radiologists (4), resident radiologists (4), medical students (3/year x 4 years = 12) and novices (3). We will, in the course of our discussion, refer to the members of the first two categories put together as *doctors* and the last two as *non-doctors*. All the radiologists were employed as radiologists at CARE multispeciality hospital, Hyderabad, India. The staff radiologists have an experience ranging from 11 years to 58 years (*M* = 30.75). Resident radiologists had at least 3 years of training.

3.2 Experimental Procedure

We used *with-in* subjects design to study the role of contra-lateral symmetry present in chest X-rays. Thus, each observer was shown both single and double lung images. All the 33 images were shown to each participant, in random order. The task given to a participant was to report the profusion level of each zone in a written form. The profusion level of a zone refers to the concentration of small opacities in that zone. Refer to Section 1.1 explaining the profusion categorization and division of lung fields into zones, while diagnosing pneumoconiosis. Information such as localization of anomalies, their size and shape were not asked, as the profusion level categorization is of primary importance to clinical settings. We used 4 (0, 1, 2 and 3) levels of profusion categorization rather than 12 levels used in ILO standards [21]. This is to minimize the complexities involved in the ILO classification procedure.

Figure 3.2 Experimental setting showing the eye tracker and monitor used in our experiment.

An observer's gaze was tracked using remote/head free eye tracker (Model Eyelink 1000, SR research, Canada [10]). Observers could freely move their heads while viewing the chest X-rays, as in clinical settings. The approximate distance between the observer and the screen was around 60 cm. The images were displayed on a 17-inch LCD monitor with a refresh rate of 75 Hz. The mean spatial accuracy of the eye tracker used was 0.5° visual angle and the sampling rate was 500 Hz. In the present experimental settings, this roughly corresponds to a tracking error of 15 image pixels. Figure 3.2 shows the experimental setting with the eye tracker and monitor used in our experiment.

Before the beginning of experiment, all the subjects were given training wherein the concept of *profusion level* was explained using some sample chest X-rays. The procedure for dividing each lung field into zones was also explained in this training. The experimental procedure consisted of following steps:

1. As the subject enters the eye tracking room, he/she was made to fill a consent form.
2. Then the subject went through a training session where the division of lung fields and profusion level categorization are explained. Experimental procedure was also clearly explained to the subject.
3. At the start of the experiment, a cover story (Figure 3.3) was shown to the subject.
4. A 9-point camera calibration was done.
5. Experimental images were shown one after the other to the subject
    - Unlimited time was given to view each image/case.
    - After complete observation of each image, the subject had to press the 'space' button on the keyboard.
    - Then the same image in small size was shown, when the subject had to note down the profusion level for *all 6 zones* in the report form given to him/her.
    - After the subject was done with noting down the profusion levels, he/she had to press the 'space' button again to get to signal completion of this case.
    - Then a blank screen with small dot at the center appeared, for drift correction, followed by next image/case.
    - In this way, all the 33 images were shown in succession.

Eye-movement data, response times, profusion levels were recorded for each observer and for each image.

3.3 Eye movement Terminology

The points where the observers look, in chest x-rays, are called fixation points and the straight-line paths between different fixation points are called *saccades*. *Saccade maps* are the images with saccades and fixations of an observer superimposed on to the original image. Figure 3.4 shows a sample saccade map. The term *saliency* refers to the likelihood of a location to attract the eye fixations of the observers. Every point in the given image can be given a saliency value based on the actual eye fixations of the

We are interested in how people look at single and double lung images. You will be shown a set of chest films. Imagine that you are practicing physician and consider that each film belongs to one of your patients and you wish to make a diagnosis. Some of the films may be normal; others may contain single or multiple abnormalities. None of the images are intended to be tricky. Press a button on the keyboard whenever you are ready. When you are satisfied that you have seen everything on the film, push the button. Then note down the profusion level of each zone in the diagnosis forms given to you.

Figure 3.3 Cover story used in our experiment

observers. Several computational models have been proposed in the literature to predict the saliency of all the points in a given image and in a given context.

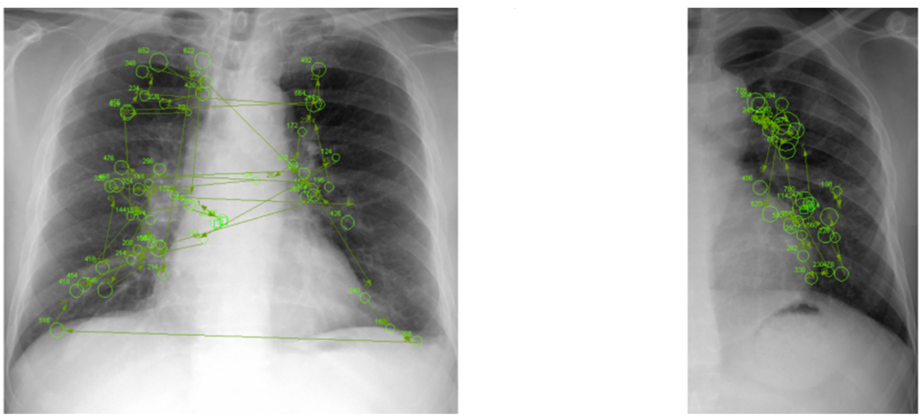


Figure 3.4 Sample saccade maps showing the eye movements of a participant, recorded while he was viewing (left) a double lung image and (right) as single lung image.

Fixation maps are the binary images with fixation points as bright pixels and all the remaining points having zero value. Saliency maps are images for which each pixel value represents the saliency at that location. *Saliency maps* overlaid onto the original image are generally referred to as Heat maps. In order to obtain a continuous saliency map of an image from the eye tracking data of a user, we convolve a Gaussian filter across the user's fixation locations. The intensity of the fixation points is directly proportional to their viewing time. Figure 3.5 shows a sample *human saliency map*, which is derived from the fixation points by Gaussian filtering.

To summarize, the basic terminology used in a general eye tracking study are listed below:

Fixation points Points on the image where observers look (fixate)

Saccades Straight line paths between different fixations

Saccade Map Image with saccades and fixations superimposed onto the original image

Saliency Likelihood of an image location to be fixated

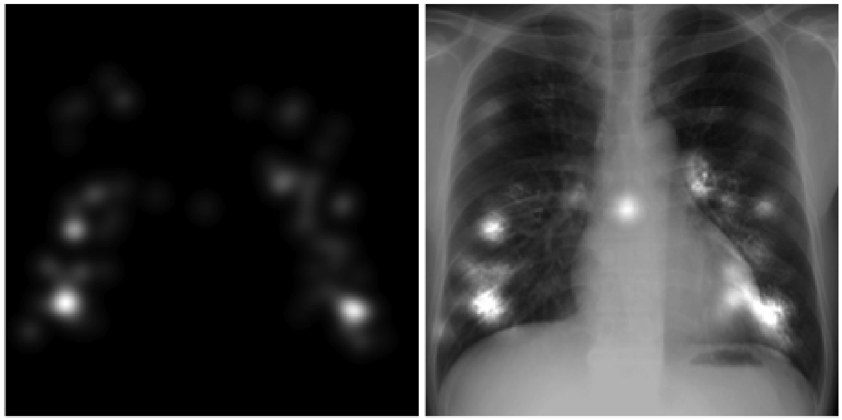


Figure 3.5 (Left) Human saliency map of an observer. (Right) Saliency map overlaid onto the original x-ray

Fixation map Binary images with fixation points as bright pixels

Saliency map Images with each pixel value representing saliency at that location

Heat map Saliency maps overlaid onto the original image

Saccade Velocity Velocity of eye movement during the saccade

Saccade Amplitude Length of the saccade

3.4 Concluding Remarks

In this chapter, we have discussed the important details and procedure concerning our eye tracking experiment, which forms the core of our study. We have also discussed some important terminology related to a general eye tracking study. In order to study the role of expertize, we used observers of different expertize levels and; to study the role of contralateral symmetry (CS), we used single and double lung images in our experiment. Eye movements of the observers were recorded, using a remote head free eye tracker, to enable the study of the role of various factors on the eye movements of the observers. We believe that we have taken all the necessary precautions, while designing our experiment, to maintain the construct, internal and external validities (see section-1.3 explaining these validities). In next chapters, we will discuss analysis of the experimental data and the corresponding results and inferences.

Chapter 4

Role of Expertize and Contralateral Symmetry

As discussed in section-2.1, expertize seemed to play a very important role in the assessment of Pneumoconiosis. To our knowledge, there are no studies on how the different levels of expertize affect the diagnosis of the disease. We also discussed, in section-2.1.1, the importance of contralateral symmetric (CS) information in diagnosing lung disease and lack of perceptual studies on the role of CS information. This chapter studies the role of expertize and CS on the assessment of Pneumoconiosis, through the analysis of eye tracking experimental data. In this chapter, we will analyze the effect of these factors on the diagnostic performance and then on the eye movements of the observers.

4.1 Analysis of Performance

Different observer studies on diagnosis of Pneumoconiosis have used different performance measures to quantify the observers diagnostic performance. Liddell et al. used information theory concept and used *information transmitted* as an inverse measure of observer error [58]. This approach counts the total number of matches and mismatches in each of the profusion level categories and used Shannons information theory concept to calculate the average amount of information transmitted. Morgan et al. used ROC analysis to analyze the observer performance [50]. For ROC analysis, the study considered the profusion rating of 0/1 or below as negative and any case given with a rating of 1/0 or greater as positive.

Some other methods of analysis such as *analysis of variance* have been used in the literature [75]. This variation in the methods of analyses used in the observer studies on Pneumoconiosis diagnosis is not unexpected given the complexity of the ILO classification procedure. Different performance measures capture different aspects of observer error. For example, the statistic of average amount of information transmitted captures total number of mismatches and matches. ROC analysis captures the discrimination capability of observers between normal and abnormal ones.

In our present study, instead of using a single measure, we used different performance measures to evaluate the observer error. *Sum of absolute differences* (described below) is taken as primary measure of observer error in our analysis as this measure is directly proportional to the observer error. Since, the *sum* *of absolute differences* do not capture various aspects of the observer performance, we have also used *penalize over* and *penalize under* to get deeper insights into the role of CS information and expertize. Given the number of observers in each group, we used non-parametric statistical tests, to study the significance of the results. Statistical tests with p-values less than 0.05 are considered significant. Two tailed p-values are considered whenever two groups are compared.

4.1.1 Sum of absolute differences

The observer error is obtained by taking the *average sum of absolute differences* between the profusion ratings, for each lung, and the ground truth profusion values as follows.

$$\text{Observer Error,} O = \frac{1}{n}\sum_{i=1}^{n}|r-p| \qquad p, r \in \{0,1,2,3\} \quad and \quad O \in [0,3]$$

Where n: Total number of Zones, p: Ground truth profusion rating, r: Observer's profusion rating

Thus, this performance measure is directly proportional to the observer error. This error is analysed by deriving the average error values for every individual and expertize group as follows: The observer error O_I for each individual is obtained by averaging O across all the images. Finally, the average error for each *expertize group*, O_G, is obtained by averaging O_I across different observers in that expertize group. O_I and O_G are shown in the top and bottom rows, respectively, in Figure 4.1 for both double and single lung images.

Some observations can be made from the plots in figure 4.1. The observer error O (grey bars) for double lung images is seen to vary significantly with expertise, which was confirmed by Kruskal-wallis test ($\chi^2(6) = 13.38, p = .038$). Thus, for double lung images, there is a decrease in error with increase in expertise; however, this trend was not present in the case of single-lung images.

It can also be showed from Wilcoxon signed rank test, that the observer error for single lung images ($Mdn = 0.813$) is significantly higher ($Z = 3.13, p < .001$) than that for double lung images ($Mdn = 0.620$). This shows that contralateral symmetry plays an important role in the diagnosis of pneumoconiosis. In order to study the role of contralateral symmetry across different expertise groups, we have studied how the difference in the observer error, between single and double lung images, varies across different groups. Mann-Whitney test showed that there is a significant difference between doctors ($Mdn = 0.38$) and non-doctors ($Mdn = 0.18$), when considering the difference of observer error between single and double lung image, with more difference in doctors than in non-doctors ($U = 28, p = .038$). As mentioned earlier, 'doctors' represent both residents and staff, and 'non-doctors' represent other groups.

Based on these results, it can be concluded that CS information plays an important role in the diagnosis of pneumoconiosis and it is perhaps used more effectively by doctors. Next, we did some finer analysis to see how CS affects the actual profusion ratings across different expertise levels.

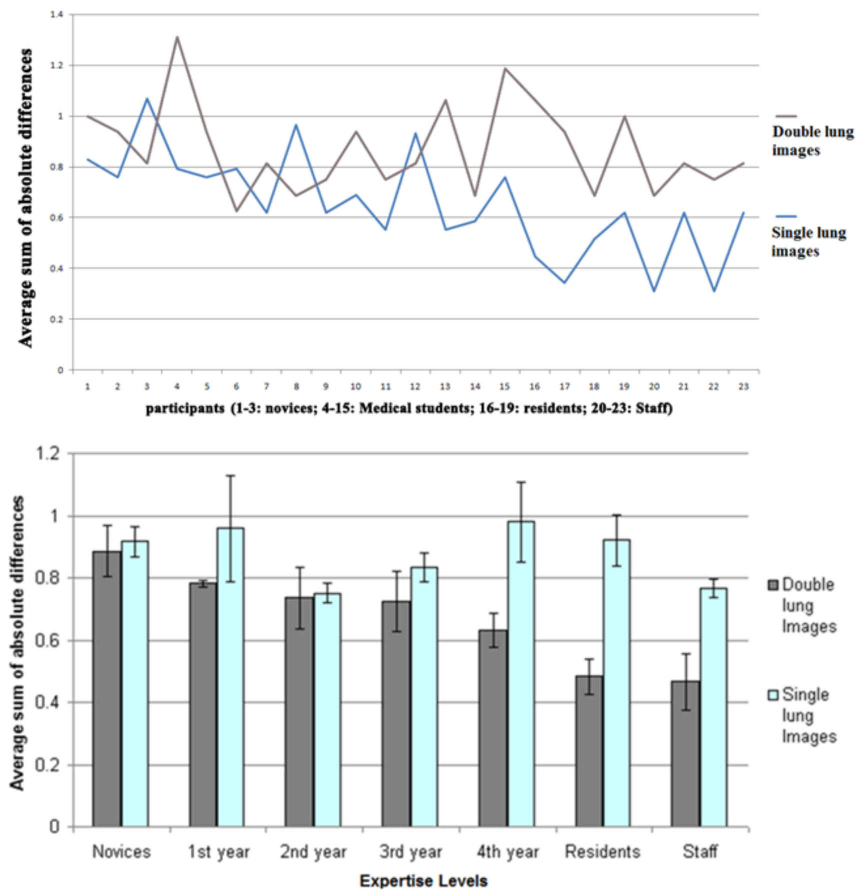


Figure 4.1 Performance of different expertise groups for both single and double lung images. Above: Line chart of the average sum of absolute differences for all the observers. Below: Average sum of absolute differences for each group.

4.1.2 Penalize Over and Penalize Under

Penalize Over/Under are the total number of times an observer has given a profusion rating *higher/lower* than that of ground truth profusion value. 'Penalize over' shows the over estimation of the profusion levels by the observers, whereas 'penalize under' reflects the under estimation of the profusion levels by the observers. Figure 4.2 shows the average penalize over and average penalize under values, for observer groups of different expertise levels.

From figure 4.2(a), it can be seen that except the staff radiologist group, observers tend to penalize over more for double rather than single lung images. More generally, when considering all the observers, the Wilcoxon singed rank test revealed that there is more penalize over ($Z = 2.13, p = .033$) in double lung images ($Mdn = 0.31$) than in single lung images ($Mdn = 0.25$). Figure 4.2(b) shows more average penalize under in single lung images than in double lung image. Wilcoxon signed rank test

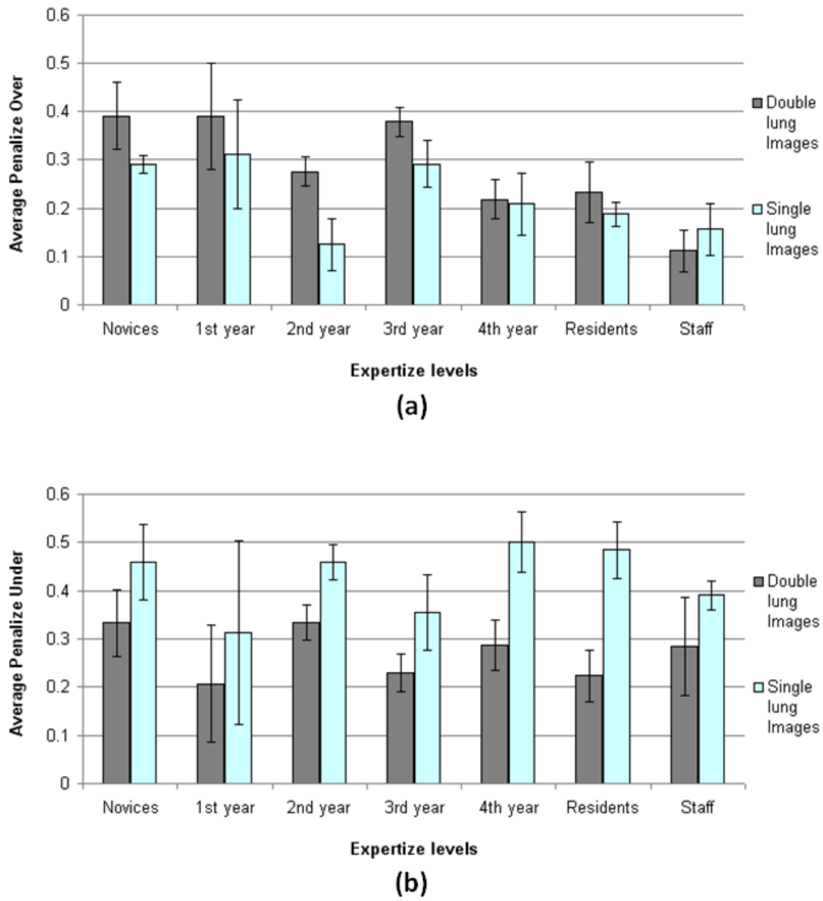


Figure 4.2 (a) Average penalize over and (b) average penalize under, for different expertise groups. These are shown for both single and double lung images

revealed that there is indeed more penalize under ($Z = 3.89, p < .001$) in single lung images ($Mdn = 0.34$) than in double lung images ($Mdn = 0.28$).

From above analysis, it appears that observers give a higher rating to a zone when the contralateral region is available for comparison. Hence, it can be concluded that CS has a role not only in correctly diagnosing pneumoconiosis but also in correctly gauging the severity i.e. assigning profusion ratings.

4.2 Time analysis

Figure 4.3 shows the average time taken for diagnosis, by different observer groups. Mann-Whitney test revealed that doctors ($Mdn = 16.69s$) took less time ($U = 21, p = .011$) than non-doctors ($Mdn = 33.24s$) in the diagnosis of double lung images. This is consistent with results of previous studies [19, 71] for other anomalies like cancer, where it was found that experienced viewers are quick and efficient when compared to non-experienced viewers.

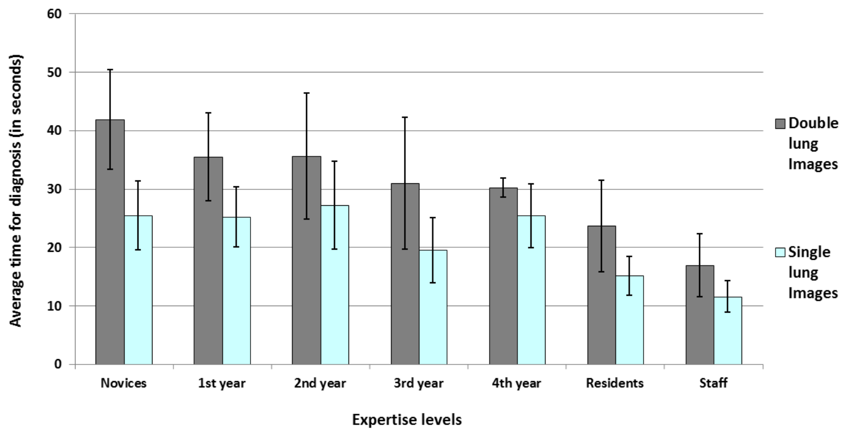


Figure 4.3 Chart showing the average time (in seconds) for diagnosis for single and double lung images and for different expertise groups

In order to assess the role of contralateral symmetry, we have compared the average time taken for double lung images with *double* the average time taken for single lung images. Wilcoxon signed rank test revealed that, on an average, the time taken for double lung images ($Mdn = 30838ms$) is less ($Z = 4.19, p < .001$) than double the time taken for single lung images ($Mdn = 38385ms$). It seems that CS information does affect the time needed for diagnosis. But, there might be some other confounding factors such as individual time preferences for viewing chest radiographs etc., which might be acting here. So, further studies are required to determine the exact role of contralateral symmetry on the time needed for diagnosis.

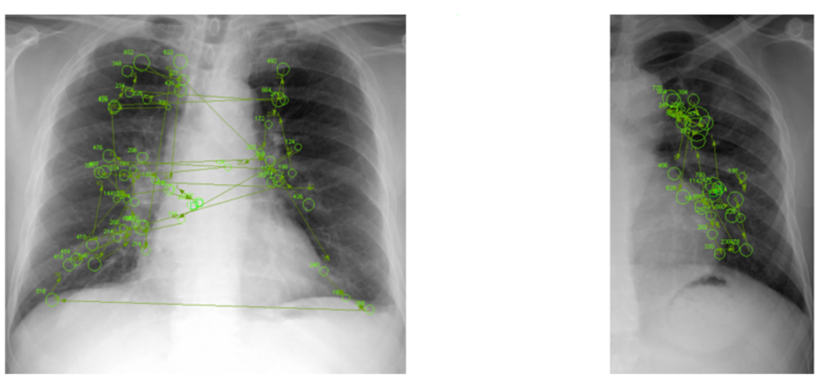


Figure 4.4 Sample saccade maps showing the eye movements of an observer, recorded while he was viewing (left) a double lung image and (right) as single lung image

4.3 Eye movement Analysis

Eye movement analysis is done at both image and zonal level. Sample saccade maps in figure 4.4 show an observer's eye movements on a double and a single lung image. At the image level, we have compared the eye movement properties such as average saccade length, average saccade velocity, average fixation time etc., across different expertise groups. Since viewing full lung images is natural in clinical settings, we have analyzed these eye movement properties in the context of double lung images only.

No significant results were found from the data of *average fixation duration*. Mann-Whitney test revealed that the *average saccade velocity* of doctors ($Mdn = 142$) is significantly higher ($U = 20, p = .01$) than that of non-doctors ($Mdn = 123$). It has also been showed, by Mann-Whitney test, that the *average saccade amplitude* is also higher ($U = 24, p = .022$) for doctors ($Mdn = 4.68$) than that of non-doctors ($Mdn = 3.98$). Thus, doctors seem to be moving eyes more quickly and over more distances, when compared to that of non-doctors. No significant correlation has been found between the performance (sum of absolute differences) of the observers and any of the above eye movement properties.

4.3.1 Zonal Eye movement Analysis

Figure 4.5 shows *average percentage of fixation times* in the different zones of double and single lung images, for different observer groups. In all images (single and double lung), Wilcoxon signed-rank test revealed that observers preference order for zones appears to be middle, lower, upper: middle zones ($Mdn = 0.21$) ($Z = 5.67, p < .001$) than lower zones ($Mdn = 0.13$), ($Z = 3.37, p < .001$) than upper zones ($Mdn = 0.10$). This might be attributed to the fact that the middle zone has more parenchyma than other zones.

Table 4.1 Average percentage of fixation time in lung zones with different profusion ratings given by the observers

Profusion Ratings	Avg. percentage fixation time			
	Novices	**Med. Students**	**Residents**	**Staff**
(double lung images)				
0	0.16	0.21	0.21	0.27
1	0.27	0.27	0.28	0.3
2	0.28	0.27	0.27	0.25
3	0.29	0.25	0.24	0.18
(single lung images)				
0	0.17	0.21	0.22	0.23
1	0.29	0.26	0.27	0.25
2	0.33	0.28	0.28	0.28
3	0.2	0.26	0.22	0.24

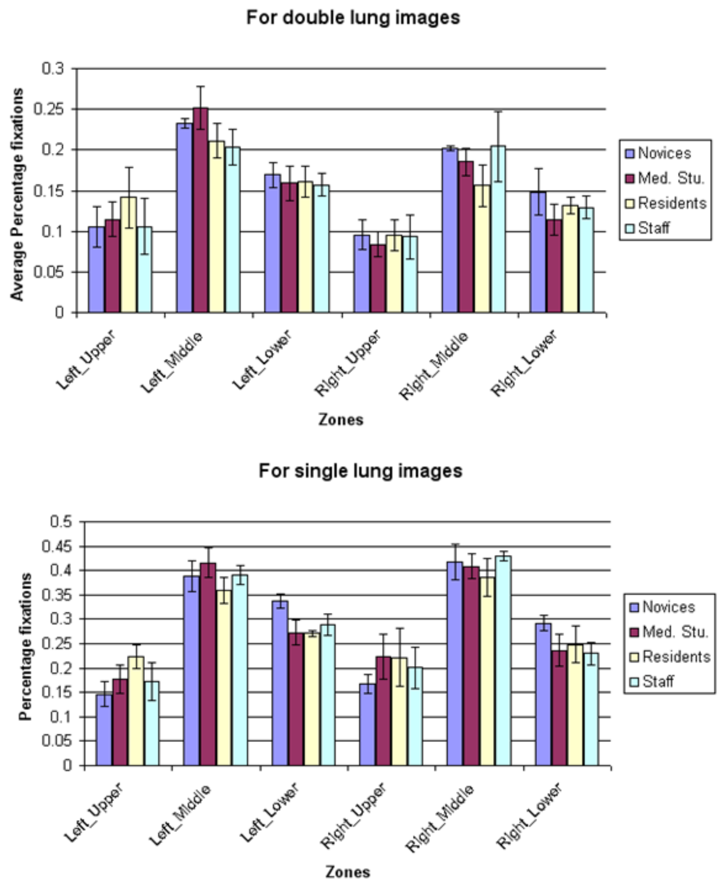


Figure 4.5 Average percentage of fixation times in different zones of (Above) double lung images and (Below) single lung images, for different observer groups

Table 4.1 shows the average percentage fixation times in lung zones with different profusion ratings given by the observers. Mann-Whitney test showed that the average percentage fixation time in lung zones with observer ratings of 1 and 2 is significantly higher than in the lung zones with observer ratings of 0 and 3, in both single lung images ($U = 636, p < .001$) and in double lung images ($U = 556, p < .001$). This shows that, whether the image is of single lung or double lung, zones considered by the observer as definite normal (profusion rating - 0) and definite abnormal (3), are less viewed when compared to that of other zones.

Table-4.2 shows the average percentage fixation times in lung zones with different absolute observer errors (absolute difference between the observer rating and the ground truth profusion value). In case of double lung images, Wilcoxon signed rank test showed that less time is spent on those zones with absolute observer error of 3 when compared to that of other zones with absolute errors of $0(Z = 3.13, p = .001)$, $1(Z = 3.65, p < .001)$ and $2(Z = 3.95, p < .001)$. No significant differences of the fixation times are seen between the zones with absolute errors of 0, 1 and 2.

Table 4.2 Average percentage of fixation time in lung zones with different absolute observer errors

Absolute Error	Avg. percentage fixation time: Novices	Med. Students	Residents	Staff
(double lung images)				
0	0.25	0.26	0.25	0.26
1	0.27	0.27	0.26	0.32
2	0.26	0.28	0.3	0.27
3	0.21	0.18	0.19	0.15
(single lung images)				
0	0.35	0.31	0.28	0.32
1	0.33	0.28	0.3	0.37
2	0.21	0.25	0.26	0.27
3	0.11	0.16	0.17	0.03

In the case of single lung images, Wilcoxon signed rank test revealed that zones with absolute error of 3 are less ($Z = 3.38, p < .001$) viewed than zones with absolute error of 2 which are in turn viewed less that zones with absolute errors of both 0 ($Z = 2.68, p = .006$) and 1 ($Z = 2.80, p = .003$). No significant difference of fixation times is seen between the zones with absolute errors of 0 and 1.

4.3.2 Gaze Transitions vs. Performance

'Gaze transitions' refer to the average number of saccades with their initial position in the left lung region and their final position in the right lung region or vice versa. Only those saccades with difference of less than 50 pixels in y-coordinates of the initial and final fixations are considered. These gaze transitions gives an approximate measure of comparisons made by the observer, between left and right lung regions, by the foveal vision.

Analysis showed a strong correlation (Person's Correlation coefficient, $r = -0.953, p = .047$) between gaze transitions and the observer error, in the case of resident radiologists. No such correlation was found in the case of staff radiologists. This shows the importance of the role of contralateral symmetry in the case of resident radiologists. Even though, CS seems to play an important role in the case of staff also, there is no significant correlation between their gaze transitions and observer error.

4.4 Discussion

The aim of the present study is two-fold. On the one hand, it involves the study of the role of CS on the diagnosis of pneumoconiosis and its influence on the readers with different expertise levels. On the other hand, it involves the study of eye movements of readers with different expertise levels.

4.4.1 On the Role of Contralateral Symmetry

Analysis of observer error (*average sum of absolute differences*) indicated that CS plays a significant role in the diagnosis of pneumoconiosis and its role is more important in the case of doctors than in the case of non-doctors. This shows that training and experience plays a very important role in learning *how to* use the contralateral symmetric information present in the chest radiographs.

The analysis of *penalize over* and *penalize under* gave some indications for how the CS information present in the chest radiographs is helping in correctly estimating the profusion in a zone. Analysis showed that observers give more rating to a zone when they have other side of lung to compare with, than not. Giving more rating may have both positive and negative effects, depending upon the true profusion value. One of the previous studies [77] showed that application of ILO classification could result in roentgenographic underestimation of asbestosis (a variant of pneumoconiosis). Our results also showed that there is a general tendency to give less profusion rating, in the case of staff radiologists and, our analysis showed that this tendency increases when there is no CS information available to the reader, leading to more observer error. Thus, CS information present in the chest radiographs helps in diagnosing the pneumoconiosis by reducing the tendency of giving less profusion ratings. In other words, CS helps reduce tendency in conservative judgment.

Even though our study showed that CS information present in the chest X-rays helps in diagnosing pneumoconiosis, we still do not know how this CS information is helping in diagnosis. In more precise terms, we still do not know, at what level this CS information is useful i.e. at image level or zonal level or at intra-rib region level. At this point, we cannot make any conclusions regarding this. More experiments are needed to study these more detailed aspects of contralateral symmetry and how these are affecting the diagnosis.

4.4.2 On Eye movement analysis

Analysis of eye movements indicated that, in both single and double lung images, middle zones are most viewed. Further, doctors move their eyes more quickly and over large distances when compared with non-doctors. This is a reasonable result to expect given the fact that doctors have considerably more training and experience, when compared to non-doctors.

Analysis showed that zones where the observers made high absolute error (3) are zones where less time was spent in viewing. This indicates the importance of time in the diagnosis of pneumoconiosis. A clear decrease in fixation time, with increase in observer error (except in the case of 0 and 1), in single lung images, suggests that time has a very important role to play in single lung images. Less fixation time in the zones which observers think as clearly normal and clearly abnormal, and less fixation time in the zones of high observer error indicates that zones need to be looked at more carefully even when the observer thinks it as clearly normal or abnormal. In other words, for better diagnostic results, all the zones should be looked at carefully i.e. X-rays should not be speed-read.

4.5 Concluding Remarks

This chapter studied the role of expertise and contra-lateral Symmetry in the diagnosis of Pneumoconiosis via a gaze tracking experiment. Results indicated that Expertise and CS play important roles in the diagnosis of pneumoconiosis. A key finding of our study is that the presence of CS information alone does not help improve diagnosis as much as *learning how to use* the information. This learning appears to be gained from focused training and years of experience. Hence, good training for radiologists and careful observation of each lung zone may improve the quality of diagnostic results.

It has been shown that CS information plays a more important role in the case of residents and staff. For residents, the eye scanning strategies seem to play an important role in using the CS information present in chest radiographs; however, in staff radiologists, peripheral vision or higher level cognitive processes seems to play a role in using the CS information.

Further experiments are required to determine the exact role of CS in diagnosing chest radiographs i.e. how exactly this information is being used by the radiologists. Since our experiment involves only the chest radiographs of pneumoconiosis, which is a diffused lung disease, the results may not be applicable to localized lung diseases such as lung cancer etc. These issues remain the topic of future work.

Chapter 5

What attracts the observer's eyes while reading chest x-rays of pneumoconiosis

In the last chapter, we have addressed the first two (of the six) research questions stated in section-1.3: i) the role of expertize and CS on various behavioral aspects of the observer like diagnostic error, time and eye movements of the observer. ii) the changes in the distribution of eye fixations with observer error and observer assessment of pneumoconiosis. In this chapter, we address the remaining four research questions. Our aim is to get some insights into the factors guiding the attention of the observers with different expertize levels. We mainly concentrate on the study of the role of anatomical features and bottom-up saliency in guiding the fixations of the observers. Section-5.1 gives basic background and some important points regarding data analysis. One of the basic assumptions, behind these perception studies, is the existence of common factors in guiding the attention of different observers. In section-5.2, we will validate this assumption by studying inter and intra observer consistency of eye fixations and in section-5.3, we study the role of anatomical features and bottom-up saliency in guiding the eye fixations. Finally, in section-5.4, we will study how the role of these different factors, changes with viewing time.

5.1 Introduction

Understanding the perceptual and cognitive factors impacting the reading of chest x-rays, helps in developing better image acquisition systems, better training regimen for radiologists and development of better computer aided diagnostic (CAD) systems [45, 60]. Several studies have been done in the past, trying to understand the cognitive and perceptual factors underlying the reading of medical images, through the careful study of eye movements. Kundel et al. [54], based on eye movement studies, suggested the holistic perception theory for the mammogram interpretation. An eye tracking study on CT images [62] suggested that this theory of holistic perception might not hold in the case of brain CT images. The perceptual and cognitive factors underlying the reading of medical images depend on several factors such as the type of image, task for the observer etc. There is no single cognitive theory that can explain the behavioral data of all the radiologists reading different types of medical images in different settings.

Human eye movement behavior depends on both bottom-up mechanisms (sensory-input) [42] and higher order top-down mechanisms (knowledge) [94]. Bottom up mechanisms are mainly stimulus driven whereas top-down mechanisms are mainly task driven. These two mechanisms together guide the eye movements of the observers. The individual role of these mechanisms depend on the stimulus, given task, expertize and other cognitive abilities of the observer. For example, while free viewing natural scenes, bottom-up mechanisms play more role than top-down mechanisms in guiding the gaze of an observer. On the other hand, if the task is finding a particular object in the scene, top-down mechanisms play more important role [94]. The interaction between the bottom-up and top-down mechanisms is still not understood completely.

Since interpreting medical images is highly task dependent, it is generally expected that top-down mechanisms play a very important and significant role in guiding the observers attention, whereas bottom-up processes might not play an important role. But, a recent study on brain CT images [62] showed that bottom-up mechanisms also play a significant role in guiding the eye movements of neurologists looking for stroke lesions on brain CT images.

Almost all the perception studies in the past have been done in settings where the observer has to search for some localized abnormality [54, 52, 51, 19, 22, 71] like that of lung tumors. Pneumoconiosis, unlike localized lung diseases, is diffused and perceptual factors underlying the assessment of pneumoconiosis might be quite different from that of searching for localized lung diseases.

5.1.1 Data Analysis

From the eye movement data, we have discarded the first fixation point on each image, to reduce the bias introduced by the central fixation marker, which is shown before each image for drift correction. We found that there is a large variability in the number of fixations across different expertize groups since we gave unlimited time to view the images. The mean number of fixations on an x-ray image for all the observers is 79.77 ($\sigma = 39.62$). This value being 113.02 for novices, 87.99 for medical students, 62.26 for residents and 47.66 for staff radiologists. We found (from the analysis reported in chapter-4) that the higher expertize groups are fast and efficient, whereas lower expertize groups are slow and inefficient. Some observers seemed to spend more time even after they have completely assessed the profusion ratings for all the zones, to crosscheck their assessment. Since such fixations do not add to the importance of the underlying image features, we have considered only the first 80 fixations of an observer, for every image in our analysis.

Since most of the data did not pass the statistical test of Normality, unless otherwise mentioned, we used non-parametric tests such as Kruskal-Wallis test and Mann-Whitney tests, to study the statistical significance of the results. Statistical tests with *p-values* less than 0.05 are considered significant. Two tailed *p*-values are considered whenever two groups are compared.

5.2 Observer consistency of eye fixations

Some gaze studies [23, 13] find the image areas of interest, by analyzing the image features underlying the fixation points of radiologists. The main aim of these studies is to automatically detect the areas of interest for the radiologists, in medical images and then develop a decision support system for new trainees. A basic assumption behind these studies is that all the observers would look at similar locations in a given image. But, this assumption is not validated in any of those studies, which try to differentiate between the fixated and non-fixated locations.

We studied the consistency of eye fixations among the observers while reading chest x-rays of pneumoconiosis. In other words, we are trying to answer the question: *Do different observers fixate at same locations in a given x-ray?* If there is a reasonable consistency of eye fixations among different observers, then we can say that the fixations of an observer can be used to predict the fixations of other observers to a reasonable accuracy.

5.2.1 Inter-Observer Consistency

Kundel et al. observe [48] that, even though different observers have different scan paths, while detecting lung nodules in chest x-rays, the distribution of their eye fixations is similar. A recent eye tracking study on natural images [41] found a good consistency between the eye fixations of different observers while free viewing the natural images. There are no similar studies on the chest x-rays of pneumoconiosis or any other interstitial lung disease.

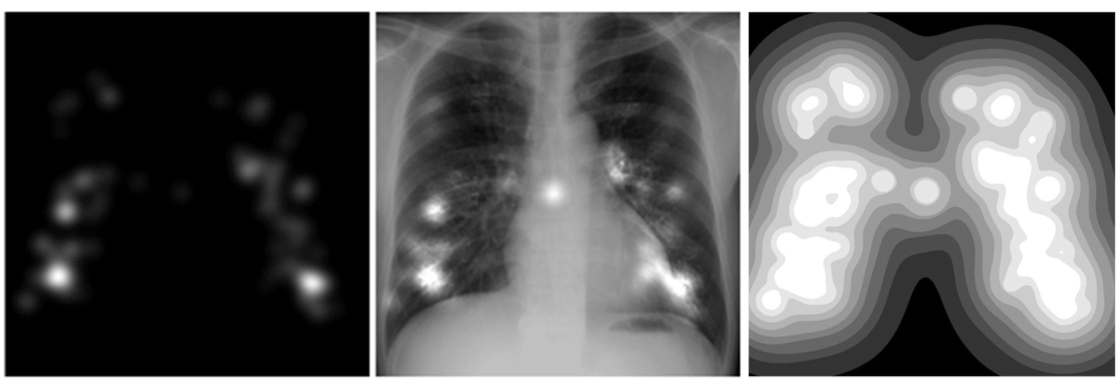


Figure 5.1 (Left) Human saliency map of an observer. (Middle) Saliency map overlaid onto the original x-ray. (Right) Soft map obtained by thresholding the saliency map to different percentage of pixels. White (brightest) pixels correspond to top 10% salient region.

We used an ROC metric to compare the fixation maps of different observers, in an expertize group. For one threshold saliency value, the *human saliency map* of one observer is treated as a binary classifier on every pixel in the image [41, 83, 32]. Saliency maps are thresholded such that a given percent of image pixels are classified as fixated and the rest are classified as not fixated. The fixations from the remaining observers, in the same expertize group, are treated as ground truth. Varying over different thresholds (figure 5.1) gives you different classifiers and thus yields an ROC curve. The area under this ROC is considered as an indication for how well the human saliency map can predict the ground truth fixations of other observers. The more the ROC area, the better is the consistency of eye fixations between the observers, in an expertize group. The ROC area would be 100 for a perfect classifier and 50 for a random classifier.

Table 5.1 AUC corresponding to different analyses, for all the 23 observers in our experiment

	NOV_1	NOV_2	NOV_3	MED1_1	MED1_2	MED1_3	MED2_1	MED2_2	MED2_3	MED3_1	MED3_2	MED3_3	MED4_1	MED4_2	MED4_3	RES_1	RES_2	RES_3	RES_4	STAFF_1	STAFF_2	STAFF_3	STAFF_4
Inter-Observer	82.6	83.1	82.5	75.6	79.8	81.3	81.8	75.6	82.6	83	82.3	73.7	76.7	79.7	73.7	77.3	79.2	78.3	79	74.8	75	74.9	75.6
Intra-Observer	80.1	83.9	83.7	76.4	82.1	84.3	86.5	78	85.4	87.1	82.9	74.1	78.5	83.2	77.9	77.8	81.8	79.7	76.5	74.4	83.1	77.4	76.2
Inter-Rib	74.7	73.9	78.6	72.3	76.8	73.1	75.8	74.9	75.1	75.8	75	67.8	66.5	72.4	68.7	77.6	72.4	75.7	75	75.4	70.5	72.8	74.7
Rib	75.4	73	79.1	72	77.4	72.3	76.5	74.1	74.6	74.8	77	68.1	66.1	72.3	69.4	77	73.9	74.3	74.4	73.8	70.2	70.6	74.1
Lung Random	76.3	75.5	76.9	72.5	73.6	74.2	74.7	74.3	75.5	76.6	77.3	68.9	68.7	74.1	70.8	75.6	72.6	75.9	72.5	73.4	71.8	74.6	75.4
GBVS	80.9	79.7	76.6	76.8	76.7	79.4	86.5	75.9	82	82.7	79.4	74.9	76.9	80.4	71	75	81.2	77.7	75.5	70.5	82.3	77.1	72.6
SIG	77	74.8	72.6	72.3	73.8	73.8	80.8	71.4	76.1	76.3	76	66.3	69.7	74.9	67.1	75.9	75.8	73.8	74.5	70.4	73.9	73.6	70.3

For each observer, and for a given image, we derive an area under the ROC (AUC) by comparing the observer's fixation data with that of other observers in the same expertize group. To reduce the systematic errors caused by the image specific data, the AUCs for all the images are averaged to get a single AUC for each observer. This AUC metric indicates the agreement of eye fixations of the observer with the remaining observers, in the same expertize group. Table 5.1 (first row) presents the AUC values for each observer corresponding to inter-observer fixation analysis.

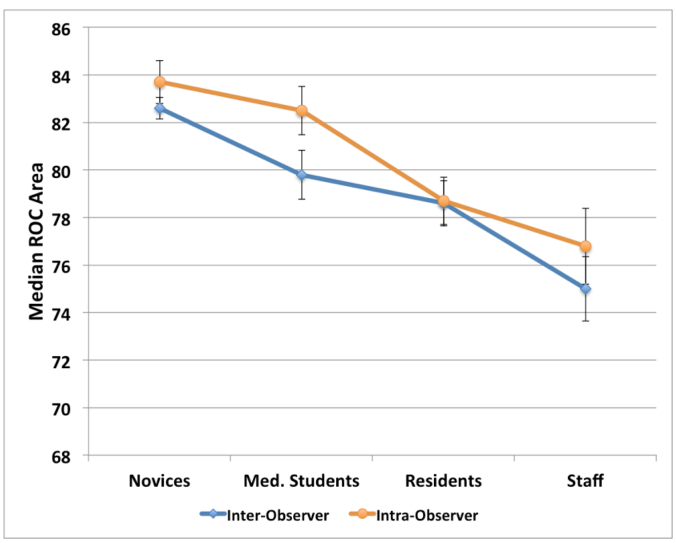


Figure 5.2 ROC Areas showing Inter-Observer and Intra-Observer consistency for different expertize groups

Figure 5.2 is a plot of the median AUC (over different observers in that group) for different expertize groups. The median AUC for all the observers is 79, which indicates reasonably good consistency of eye fixations. Hence, fixation points of an observer can be used to predict the fixation points of other observers to a reasonably good accuracy. When considered in pairs of observer classes, the difference of AUC between medical students and residents is not statistically significant (Mann Whitney U test: $U = 20.0, z = 0.485, p = 0.627$), whereas, there is a significant difference between novices and medical students ($U = 3.5, z = 2.096, p = 0.036$); and between residents and staff radiologists ($U = 0.0, z = 2.309, p = 0.021$). Thus, there is more agreement in fixations among the observers of lower expertize groups than that of higher expertize groups. Hence, it appears that there are more common factors that guide the eye movements of lower expertize groups than those of higher expertize groups.

5.2.2 Intra-Observer Consistency

Since all the images are of PA chest x-rays, we can expect some amount of intra-observer consistency as well. We can get good insights into the role of observer's reading style and also the role of anatomical features common to all the x-ray images, in guiding the observer's eye fixations. The research question here is: *Does an observer fixate at same locations, on different x-ray images?* We have not found any previous studies on intra-observer variability of eye fixations on medical images.

We used similar type of analysis, explained in the last section, to study the intra-observer consistency of eye fixations. For a given observer, the human saliency map of one image is treated as a binary classifier on every pixel in the image. The fixations of the same observer on the remaining images are treated as ground truth. Then ROC curves are drawn by the procedure explained previously and here, the AUC indicates how well the human fixation map of an observer on an image can be predict the fixation points of the same observer on other images. The AUCs for all the images are averaged to get a single AUC value for each observer. Table 5.1 (second row) shows the AUCs, for all the observers, corresponding to intra-observer analysis.

Figure 5.2 shows the median AUCs for different expertize groups. The median AUC, for all the observers, is 80.1. Thus, for a given observer, fixation map on one image can be used to predict the fixation maps on other images, with a reasonably good accuracy. Even though the AUCs seem to be decreasing with increasing expertize, the differences are not statistically significant.

5.2.3 Inter-Observer vs. Intra-Observer Consistency

From the figure 5.2, one can see that the AUCs corresponding to intra-observer analysis are marginally higher than that of inter-observer analysis, for all the expertize groups. Wilcoxon signed rank test showed the significant ($Z = 29.5, p < 0.001$) difference of AUCs between the inter-observer ($Mdn = 79$) and intra-observer ($Mdn = 80.1$) analysis. The higher AUCs for intra-observer analysis indicates that observer reading style and non-image specific features such as anatomical features are playing more important role in guiding the eye movements of the observers, than the image specific textural features. Further analysis is required to determine the role of anatomical features in guiding the eye movements.

5.3 Role of image features in predicting eye fixations

Image features can be broadly classified into three types, depending on the level of computation required to extract them. *Low-level features* like pixel intensity, colour, orientation etc., which can be directly computed from a pixel or combination of pixel values. *Mid-level features* such as blobs, holes etc. are generally computed from low level features. And, *high-level features* include anatomical structures like ribs, heart etc. In this study, we focus on the role of low-level bottom up saliency related image features and the high level anatomical features, in guiding the eye movements of the observers. By understanding the role of these features in predicting the fixations of the observers, we can get good insights into the common factors that underlie the observers visual attention.

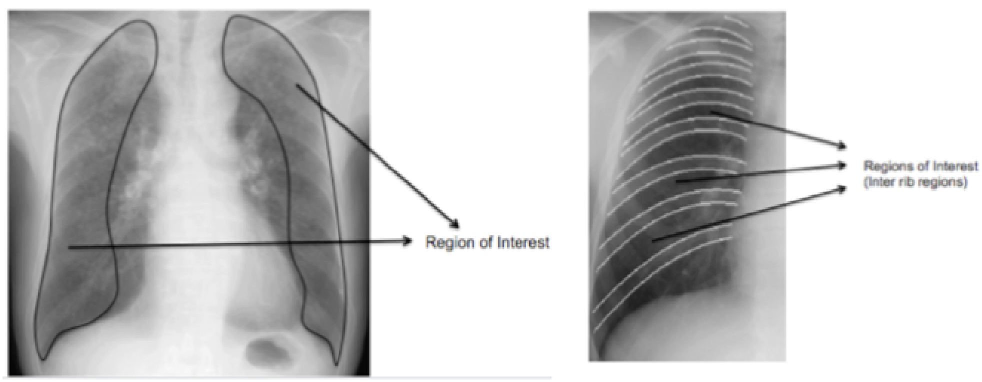


Figure 5.3 Radiographs with some marking showing the region of interests, namely lungs (left) and inter-rib regions (right)

5.3.1 Role of anatomical features

Since pneumoconiosis is a lung disease, lung regions and inter-rib regions are main region of interests (ROIs) for assessing the profusion levels (figure 5.3). Even though these ROIs need to be examined more carefully for pneumoconiosis assessment, it has not yet been established whether the observers actually look at these regions more or not compared to other regions in a chest x-ray. One of the aims of the present study is to find out which regions of chest x-ray are looked at more and thereby given more importance by the observers of different expertize level. The aim is to determine the role of anatomical features in guiding the eye movements of the observers. The anatomical features that are considered here are: lung, rib and inter-rib regions.

Anatomical distribution of eye fixations

The fixation distribution across different anatomical regions can give a rough indication of the importance given to those respective regions. To find this fixation distribution, we need to segment out those anatomical regions. We used Euler number based thresholding [92], to get roughly segmented lung

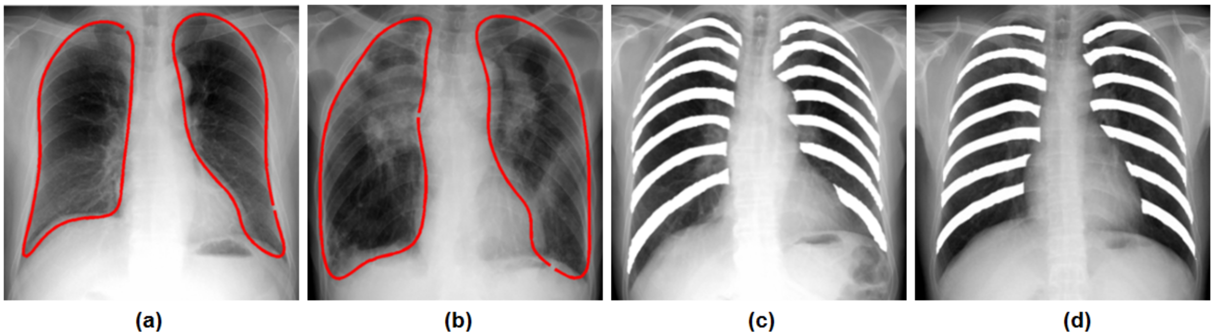


Figure 5.4 Sample (a,b) lung and (c,d) rib segmentation results

regions and then we used an active contour method, similar to the approach in [12], to finely segment the lung regions. Figure 5.4 (a,b) shows sample lung segmentation results.

Segmentation of ribs is a difficult problem. Even though it is easy to visually discriminate the ribs, it is computationally a difficult problem. The superimposition of normal anatomical structures such as thoracic vasculature, clavicles, the heart, and fatty tissue can make it hard to distinguish the edges corresponding to rib borders. We found that the current rib segmentation algorithms [59] are not good enough to clearly segment rib regions, especially in the case of images of higher disease level. Hence, a manual segmentation of rib regions was done. Figure 5.4 (c,d) presents sample manual rib segmentation results, where white pixels corresponds to segmented rib regions. The pixels which are inside lung regions and are not belonging to rib regions are considered as belonging to inter-rib regions.

Analysis showed that, on an average, around 84% of fixations are inside lung regions, for all the observers. Thus, lungs regions are given clear importance over other regions. This is expected given the nature of the task given for the observers. Given the difference in the areas of different anatomical regions, the percentage of fixations in different anatomical regions would not give good indication of the importance given to those respective regions. So, instead, we used a *fixation density* measure to find the relative importance given to a region. The fixation density of a region, for a given observer and image, can be calculated as:

Fixation density of a region

$$= \frac{(\text{Number of fixations in the region})}{(\text{Total number of fixations in the image} * \text{Area of the region})}$$

$$= \frac{(\text{percentage of fixations in the region})}{(\text{Area of the region})}$$

It should be noted that the fixation density is not equivalent to the percentage of fixations in a region and if we consider the area of entire image as 6.5536 ($256 * 256/10000$), we get the constant fixation density of 15.26 for the entire image. Table 5.2 shows the average fixation density in different anatomical regions, for all the four expertize groups, when the area of entire image is considered as 6.5536. The fixation density data, for different observers and for different anatomical regions, passed the test of

Table 5.2 Average fixation density in different anatomical regions for different expertize groups

	Total Image	Right Lung	Left Lung	All lung	Right Rib	Left Rib	All Rib regions	Right Interrib	Left Interrib	All Interrib
Novices	15.26	39.54	35.17	37.49	38.12	35.33	36.99	38.89	34.07	36.51
Med. Students	15.26	41.6	27.21	34.94	41.83	25.88	34.89	40.08	26.9	33.69
Residents	15.26	41.87	31.26	36.72	41.66	31.38	36.89	40.71	30.3	35.49
Staff	15.26	36.83	33.07	34.9	35.64	30.22	33.12	36.3	33.88	34.87
All	**15.26**	**40.54**	**29.96**	**35.56**	**40.2**	**28.82**	**35.18**	**39.4**	**29.62**	**34.58**

normality. So, we have used parametric tests, here, for statistical analysis. Here, the *right lung* refers to the anatomical right lung and it is located on the left side of the image and *left lung* refers to anatomical left lung.

Statistical analysis showed no significant relation between the expertize level and the fixation density of different anatomical regions. More importantly, there is no significant difference between the fixation density of lung regions, rib regions and Inter-rib regions. Thus, both rib and inter-rib regions appear to be almost equally fixated.

From Table 5.2, it can be seen that anatomical regions in the right lung are fixated more than the corresponding anatomical regions in the left lung region. Paired t-test ($t(390) = 12.75, p < 0.001$) showed that the fixation density in right lung region ($M = 40.54$) is significantly higher that of left lung region ($M = 29.96$). Rib regions in the right lung ($M = 40.20$) have significantly more fixation density ($t(390) = 9.575, p < 0.001$) than the rib regions in left lung ($M = 28.82$). In the same way, right inter-rib regions ($M = 39.40$) have significantly more fixation density ($t(390) = 10.351, p < 0.001$) than the left inter-rib regions ($M = 29.62$). Since, fixation density in all the lung regions considered is well above fixation density of the entire image (15.26), we can say that lung regions are highly fixated compared to the regions outside lung. This is an expected result, given that the pneumoconiosis is a lung disease and the observers have to look at only lung regions to assess the profusion ratings

The more interesting result here is that despite the theoretical importance to inter-rib regions while assessing pneumoconiosis, both inter-rib and rib regions are fixated almost equally. It is interesting to see that the right lung regions are fixated more compared to that of left lung regions. This might be because of normal reading style of the observers or the right lung is used as a reference lung against which the left lung region is compared. More experimentation is required to find the reason for this right lung dominance.

Analysis with anatomical saliency maps

A ROC analysis, similar to the study of inter and intra observer fixation consistency, was done to get more insights into the role of rib and inter-rib regions in attracting the observers gaze. This kind of analysis also helps us in comparing the relative significance of anatomical regions in predicting the observers gaze in comparison to that of human saliency maps.

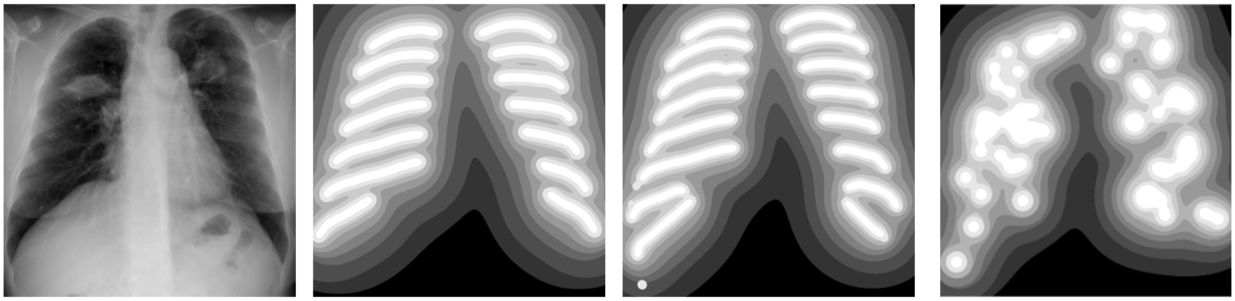


Figure 5.5 A sample x-ray image and corresponding rib, inter-rib and random lung saliency maps thresholded to different percentage of pixels

By taking fixation points along the rib and inter-rib regions, we have created rib and inter-rib saliency maps respectively. In order to analyze the importance given to points inside lung regions, we have used random lung saliency maps, which are created by taking fixation points at random locations inside lung regions and with random fixation duration. Figure 5.5 shows sample rib, inter-rib and random lung saliency maps, thresholded to different percentage of pixels.

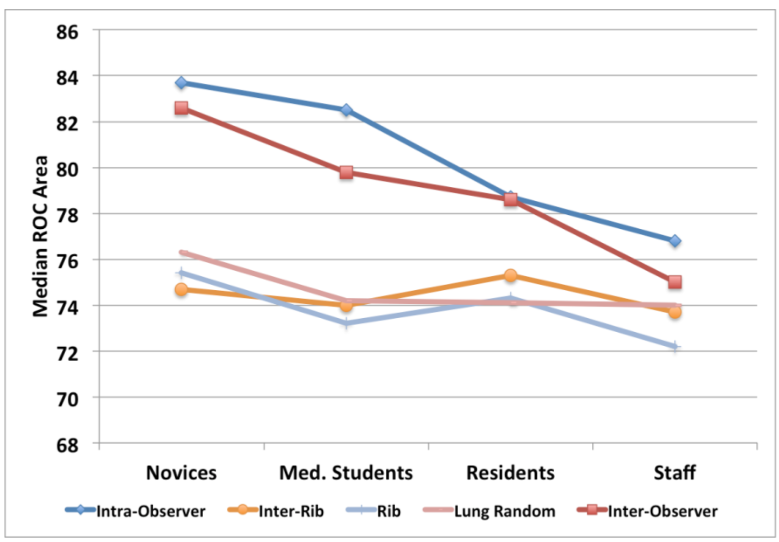


Figure 5.6 ROC Areas indicating the role of different anatomical saliency maps in predicting the eye fixations

By considering these saliency maps as classifiers and the observers' fixations as ground truth, ROC curves were derived. These curves help indicate the role of corresponding anatomical regions in attracting the observer's gaze. Table 5.1 shows the AUC values for all the observers and figure 5.6 plots the median AUCs for the anatomical saliency maps along with the AUCs for intra observer and inter-observer analysis.

Wilcoxon signed rank tests showed no significant differences of AUC between those related to rib, inter-rib and random lung saliency maps. In addition, there is a good correlation of AUCs (performance) related to rib, inter-rib and random lung saliency maps. For all the observers, Spearman's rank order correlation showed a good correlation between the performance of rib and inter-rib saliency maps ($r_s(21) = 0.907, p < 0.001$); between the performance of inter-rib and random lung saliency maps ($r_s(21) = 0.704, p < 0.001$) and; between the performance of rib and random lung saliency maps ($r_s(21) = 0.749, p < 0.001$). Thus, both rib and inter-rib saliency maps have same prediction accuracy as that of random lung saliency maps, in predicting the human fixations. There is no significant difference in the performances of these anatomical saliency maps across different expertize groups.

Comparing the performance of random lung saliency maps to that of intra-observer consistency, for all the observers, Wilcoxon signed rank test showed that the ROC areas for intra-observer saliency maps ($Mdn = 80.1$) are significantly higher ($Z = 0.0, p < 0.001$) than that of random lung saliency maps ($Mdn = 74.3$). Also, the ROC areas for inter-observer saliency maps ($Mdn = 79.0$) are also significantly higher ($Z = 0.0, p < 0.001$) than that of random lung saliency maps.

From the above results, we can say that the points on rib and inter-rib regions are given equal importance and the random lung saliency maps account for about 94% (Meidan AUC of 74.3) of the 79.0 AUC observed for the performance of inter observer human saliency maps. Thus, much of the intra and inter observer consistency can be explained by the observers giving high importance to lung regions. Still, there is a significant AUC difference between random lung saliency maps and that of inter observer human saliency maps indicating that there are common factors other than the bias towards lung regions that are guiding the eye movements of the different observers. Further analysis is required to get insights into these factors.

5.3.2 Role of bottom up saliency

Next, we examine the role of bottom up saliency of chest x-rays in guiding the visual attention of the observers. Bottom up saliency corresponds to the visually conspicuous image areas and is independent of the task given to the observer and also independent of the observer's expertize. Several studies have shown the importance of bottom up saliency in guiding the visual attention of observer while viewing natural images [42, 32, 40]. A recent study also showed that bottom up saliency plays a significant role in the neurologists viewing brain CT images [62].

Several computational bottom-up saliency models have been proposed in the literature. However, there is no single model that performs well on all types of images. Hence, we have considered 10 state-of-art bottom-up saliency models, to analyze their performance in predicting the fixations of the observers in our experiment: Itti-Koch saliency model (**IK**) [40], Graph based visual saliency (**GBVS**) [32], Image signature (**SIG**) [35], spectral residual approach (**SR**) [36], Dynamic visual attention (**DVA**) [37], Adaptive whitening saliency (**AWS**) [26]], Attention based on information maximization (**AIM**) [17], Saliency based on local self-resemblance (**SDSRL**) [79], Saliency based on global self-resemblance (**SDSRG**) [79] and context aware saliency (**CA**) [29]. Matlab codes for all these saliency models are available on their respective author's webpages [7, 4, 5, 9, 3, 6, 8, 2, 1]. We used the default parameter settings used by the authors for extracting the saliency maps of the x-ray im-

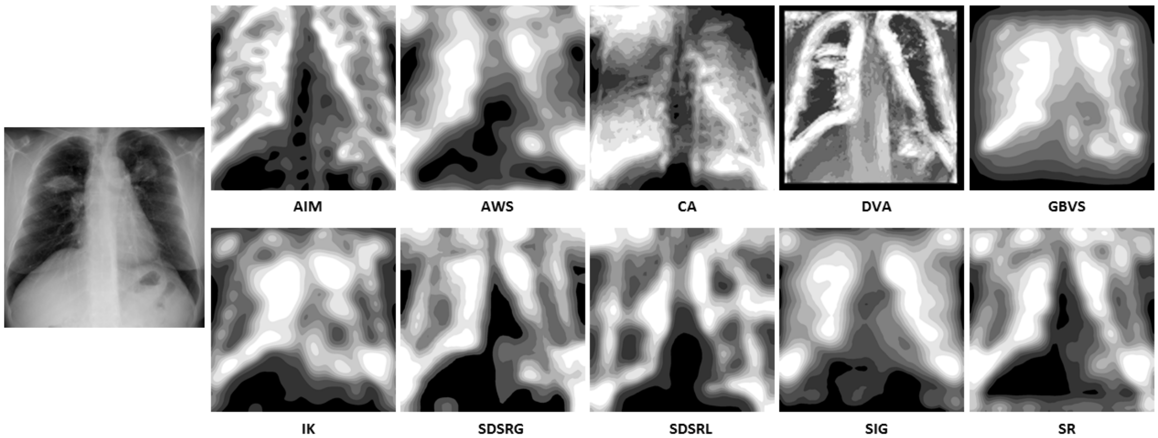


Figure 5.7 Original image and corresponding saliency outputs of various computational saliency models. The output is obtained after thresholding to different percentage of pixels

ages in our dataset. Figure 5.7 shows the sample thresholded (to different percentage of pixels) saliency maps extracted by running these saliency models on an image.

A ROC analysis was also conducted. The saliency maps extracted using these saliency models are considered as classifiers and the human fixations constitutes the ground truth.

Figure 5.8(a) shows the AUCs corresponding to different saliency models, for all the 23 observers in our experiment. It can be clearly seen from the plots that the GBVS and SIG models outperform other bottom-up saliency models. Table 5.1 shows the AUCs corresponding to GBVS and SIG saliency models, for all the observers. Wilcoxon signed rank test showed that the AUCs corresponding to GBVS saliency maps ($Mdn = 77.1$) are significantly higher ($Z = 2.0, p < 0.001$) than those corresponding to SIG saliency maps ($Mdn = 73.8$), when all the observers are considered. Even when the individual expertize groups are considered, GBVS saliency maps outperformed SIG saliency maps in predicting the fixations of the observers.

What distinguishes GBVS and IK from other saliency models is that, by design, they have more biological basis and they are based on the low level image features such as intensity and orientation. The second best model, SIG works in spectral domain. Both GBVS and IK use the same set of basic image features at multiple scales to compute saliency maps. The basic difference between them lies in how they combine activation maps at multiple scales to get the final saliency map (*combination*) and how they do *normalization*. Compared to IK, GBVS model makes long-range pixel comparison of feature values to compute final saliency maps and has more center bias [32]. It is reported [32] that GBVS outperforms IK model in predicting the saliency of observers while viewing natural images as well.

Figure 5.8(b) shows the plot of AUCs, for different expertize groups. A median AUC of 77.1 suggests that GBVS saliency model can be used to a reasonably good accuracy to predict the fixations of the observers. For all the observers (across expertize groups), bottom up saliency plays an important role in

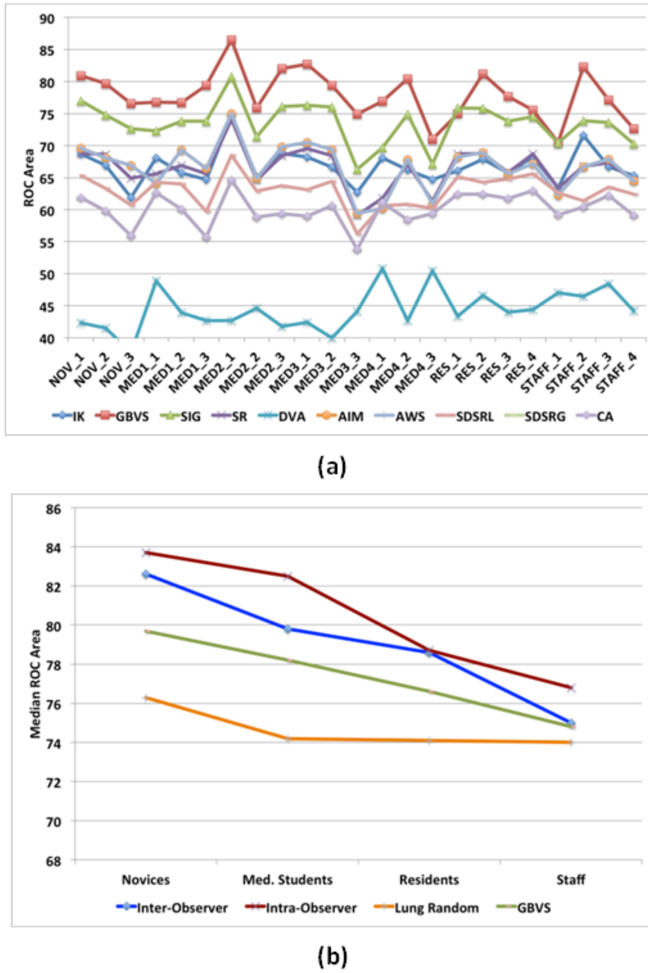


Figure 5.8 ROC Areas corresponding to (a) different saliency models for all the observers; (b) Inter-observer, intra-observer, random lung and GBVS saliency maps for all four expertize groups.

guiding the eye fixations. Even though the median AUCs of lower expertize groups are higher than that of higher expertize groups, the differences are not statistically significant.

Comparisons with the performance of inter, intra observer and random lung saliency maps

Given the above conclusion it is pertinent to ask: What is the role of the bottom-up saliency in the observed inter-observer and intra-observer consistency? Figure 5.8(b) shows the AUCs for GBVS saliency maps along with inter-observer, intra-observer and random lung saliency maps, for different expertize groups.

For all the observers, the AUCs for GBVS saliency maps are significantly higher ($Z = 22.0, p < 0.001$) than those of random lung saliency maps. Also, the AUCs for intra-observer saliency maps are significantly higher ($Z = 13.0, p < 0.001$) than those of GBVS saliency maps. But, the difference between GBVS and inter-observer saliency map ROC areas is not significant ($Z = 92.0, p = 0.162$), when considering all the observers.

From the above results, it is clear that the GBVS saliency model performs significantly better than random lung saliency maps in predicting the eye fixations. We can also say that GBVS saliency model explains much of the observed inter-observer consistency. Thus, bottom up saliency is an important factor in guiding the eye fixations of observers. Even though, not significantly higher, the AUCs related to inter observer fixation consistency are marginally higher than that of AUCs related to GBVS saliency maps (Table 5.1). So, there still seems to be some common factors, other than bottom up saliency that guide the fixations of different observers.

5.4 Effects of Time

We also studied how the role of different factors, studied in the previous sections change with time. In order to study the effect of time, we did ROC analyses by taking different number of time ordered fixations starting from 5 fixations to 105 fixations. The research question we are trying to address here is: *How do the inter-observer fixation consistency, intra-observer fixation consistency, role of lung regions and role of bottom up (GBVS) saliency change with time (number of fixations)?* Analysis of this kind would give more insights into the visual strategies used by the observers in assessing the disease level of pneumoconiosis.

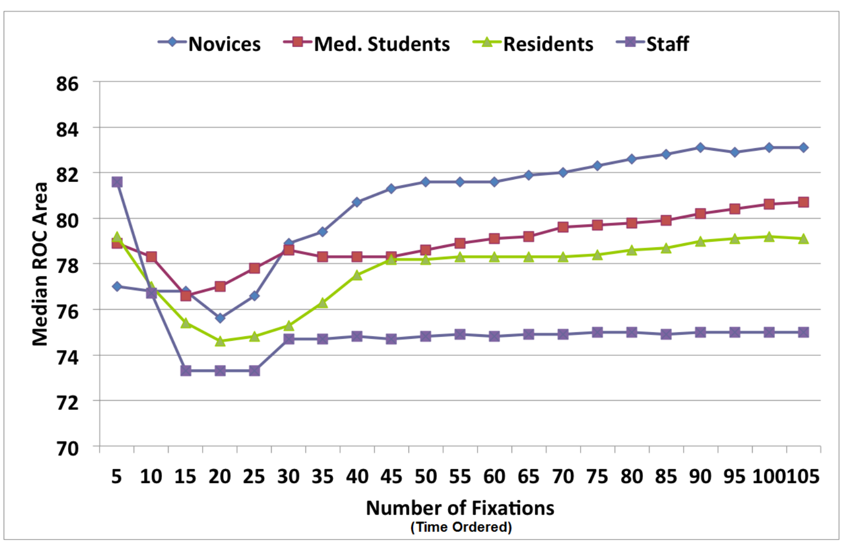


Figure 5.9 ROC areas, indicating the inter-observer fixation consistency, calculated for different fixation numbers, for all the four expertize groups

5.4.1 Inter-Observer fixation consistency

Figure 5.9 shows the AUCs, indicating the inter-observer fixation consistency, for different number of time ordered fixations and for different expertize groups. In the figure 5.9, the values of 5,10,15,... on horizontal axis correspond to first 5 fixations, first 10 fixations and so on. The inter-observer AUCs are calculated as explained previously (see section-5.2.1), by considering different number of time ordered fixations, at a time. From the AUCs in Figure 5.9, it can be seen that the inter-observer fixation consistency is initially high, which decreases to about 20 fixations and then increases slowly to plateau off. The inter-observer consistency for *doctors* (residents and staff) exhibits a different trend compared to that of *non-doctors* (novices and med. students). For doctors, the initial decrease of AUC is more rapid compared to that of non-doctors. Further, the increase of AUC after 20 fixations extends to more number of fixations in non-doctors than in the case of doctors.

Thus, different doctors appear to inspect similar locations in an image, indicating the use of similar visual strategy, for the first 5-10 fixations. After this, the rapid decrease in AUCs till 20 fixations indicates the divergence of visual strategies, across different *doctors*. Thus, we can say that the initial fixations are playing an important role in determining the later visual strategies employed by the doctors. It seems that the information gained during the first few fixations by the visual system is playing an important role in choosing the later viewing strategy.

In the case of non-doctors, the AUC changes are not rapid but the increase after 20 fixations is extending to more number of fixations as compared to that of doctors. This might be due to the reason that the lower expertize groups have more number of fixations per image than higher expertize groups. One of the reasons for the slow increase of ROC areas, after initial 15-20 fixations, might be that even though different observers are using different viewing strategies after initial few fixations, they might be covering the same areas as other observers but in different order.

5.4.2 Intra-Observer fixation consistency

In order to study, how the intra-observer fixation consistency changes with time, we calculated the AUCs (as explained in section-5.2.2), using different number of time ordered fixations. Figure 5.10 shows the chart of these AUCs for different expertize groups.

Similar to the case of inter-observer analysis, we find different trends for doctors and non-doctors. For non-doctors, the intra-observer fixation consistency decreases slowly to about 20-25 fixations and then increases slowly as fixation number increases. For doctors, the intra-observer fixation consistency decreases rapidly to about 20 fixations and then increases slowly and marginally.

Thus, a doctor seems to be following similar viewing strategy, during the first few fixations, while viewing different x-rays. His strategies seem to diverge after the initial few fixations accounting for less intra-observer fixation consistency after initial fixations. In the case of non-doctors, the less differences in AUCs across different number of fixations indicate that their viewing strategies remain the same overtime, for different images.

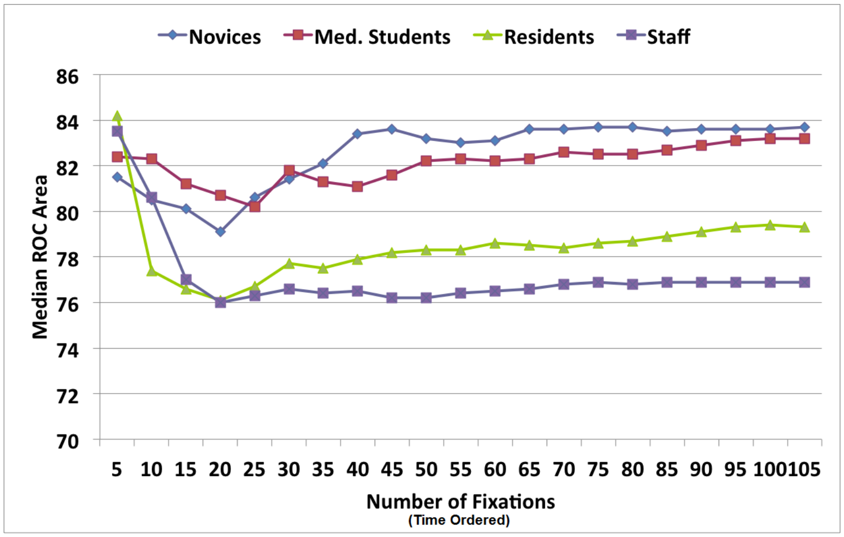


Figure 5.10 ROC areas, indicating the intra-observer fixation consistency, calculated for different fixation numbers, for all the four expertize groups

5.4.3 Role of random lung saliency maps

Figure 5.11 shows the AUCs indicating the role of random lung saliency maps in predicting the eye fixations of the observers, of different expertize groups, for different number of fixations. AUCs are calculated using the *random lung saliency maps*, as explained in section-5.3.1. This AUC reflects the importance given to the lung regions, by the observers.

Figure 5.11 shows no clear trend in the AUCs as the fixation number increase, except in the case of novices, who give less preference to the lung regions during the first few fixations. In other words, there is no significant increase or decrease in the AUCs. We can only see the marginal increase of AUCs in the case of medical students. Overall, the preference given to the lung regions remains with time.

We have done similar analysis using *inter-rib and rib anatomical saliency maps* (defined in section-5.3.1) and obtained similar results (not shown here). Thus, it seems that the role of lung regions in attracting the gaze of the observers is consistent over entire viewing period.

5.4.4 Role of bottom up saliency

Finally, we studied how the role of bottom up saliency changes with time. Figure 5.12 shows the ROC areas indicating the performance of GBVS saliency maps in predicting the eye fixations of the observers of different expertize groups, for different time ordered fixations.

From figure 5.12, it can be seen that the AUCs for novices, medical students and residents seems to decrease as the number of fixations increases. However, for staff radiologists, the AUC seems to be almost constant for different number of fixations. Thus, except in the case of staff radiologists, the

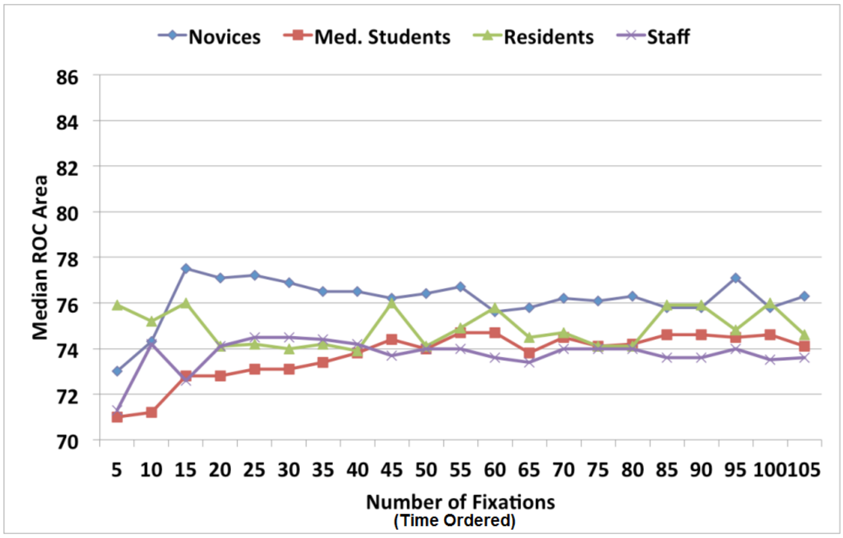


Figure 5.11 Change in Area under ROC over viewing time indicating the role of random lung saliency maps in predicting the fixations of the observers belonging to different expertize groups

bottom up saliency is playing more important role in attracting the gaze during the initial fixations when compared to later ones.

From the above analysis, it is clear that time plays an important role in determining the role of different factors underlying the eye fixations. The visual strategies used by the observers, during the first few fixations seem to be quite different from the visual strategies used later on, in their viewing.

5.5 Discussion

The study reported in this chapter can be broadly classified into three sections. First, we studied the inter-observer and intra-observer fixation consistency. Second and third being the study of role of anatomical features and bottom up saliency in attracting the fixations of the observers respectively. We analyzed the inter observer and intra-observer fixation consistency to assess the degree of fixation agreement across and within the observers. Since we found reasonably good consistency of fixations between and with-in observers, we proceeded to study the image features that distinguish the fixation and non-fixation points. In the present study, we have considered two different kinds of image features in determining their role in attracting the gaze of the observers: High-level anatomical features and low-level bottom up saliency features.

5.5.1 On Inter-observer and Intra-observer fixation consistency

Results showed reasonably good inter-observer and intra-observer fixation consistency, indicating that the saliency map of an observer can be used to a reasonable accuracy in predicting the fixations

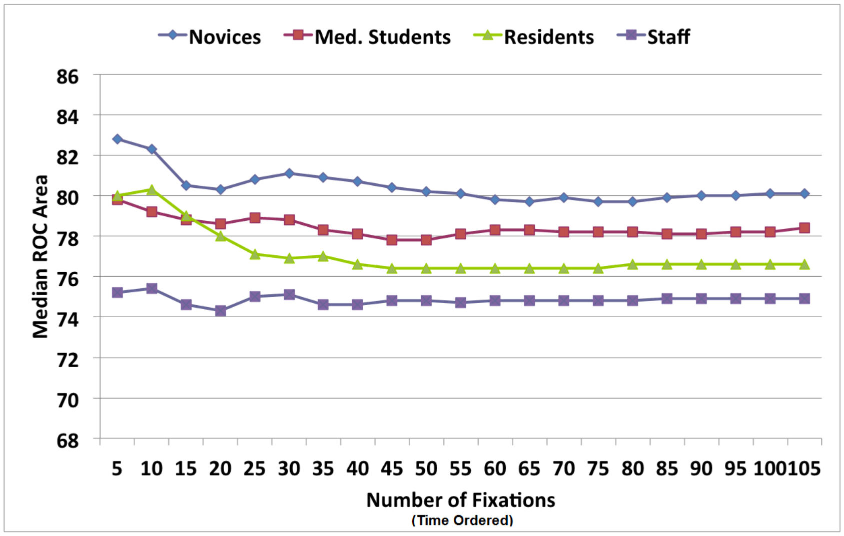


Figure 5.12 ROC Areas indicating the role of bottom up saliency (GBVS) maps in predicting the eye fixations of the observers of different expertize groups

of other observers, in the same expertize group, and also the fixations on other images, of the same observer. Significantly higher intra-observer fixation consistency than inter-observer fixation consistency suggests that the *reading style* of the observers is playing more important role than image specific features in guiding the fixations of the observers. Nevertheless, the study on the role of bottom-up saliency showed that bottom-up image features also play an important role in attracting the fixations of the observers.

The more intra-observer consistency in lower expertize groups, when compared to that of higher expertize groups shows that the fixations of the higher expertize groups are more dependent on the image content when compared to that of lower expertize groups. In the same way, higher inter-observer consistency in lower expertize groups suggest common viewing strategies in lower expertize groups compared to that of higher expertize groups. From these results, we can say that the experience is helping observers in developing their individual viewing strategy and adapt it to the image content.

Analysis with different number of time ordered fixations showed that the viewing strategies of the observers change with time. This is especially true in the case of doctors, for whom there is an initial high inter and intra observer fixation consistency, which rapidly declined till about 20 fixations. Thus the information gained during the first few fixations is playing an important role in deciding which visual strategy to use by the visual system to further read the given chest x-ray. Analysis suggested more diverse visual strategies for doctors compared to that of non-doctors.

The slow increase of the inter-observer and intra-observer consistency, after about 20 fixations might be due to different observers viewing the same regions in a given image in different pattern/order. This is consistent with the result in [48] where it was mentioned that even though different observers use different viewing patterns like zigzag, circumferential etc., they would look at the similar areas in a given x-ray. These results suggest that the locations where the observers have fixated are more important than the order in which they have fixated at these different locations.

5.5.2 On the role of anatomical regions

Pneumoconiosis, being a diffused lung disease, it is generally expected that the observers will spend more time in the lung regions compared to non-lung regions. It is also reasonable to expect more fixations in the inter-rib regions when compared to that of rib regions, as inter-rib regions carry more disease related information. Even though the first expectation of importance to lung regions is supported by the experimental data, the latter expectation is not. Results suggested that the observer gave equal importance to inter-rib and rib regions. Points on rib and inter-rib regions are considered as any other random point inside lung region. Same is the case with all four expertize groups. This is an interesting result as it suggests that information on the rib regions is also playing a very important role in the decision making process of the observers.

The fact that the inter rib and rib regions are given equal importance as any other random point inside lung region may not imply that all the points inside lung region are given equal importance. For example, it might be the case that the regions closer to mediastinum or heart may be given more importance than other regions. Further analysis is required to determine the existence of any such behavior. Another interesting result is that the anatomical right lung regions are fixated more compared to the anatomical left lung regions. This might be due to the normal viewing style of the observers or might be due to the presence of heart in the anatomical left side. Further experiments are required to get insights into this peculiar behavior. A gaze tracking experimental study showing normal chest x-rays and left to right flipped x-rays to the observers, would give good insights into such eye fixation behavior. This remains the part of future work.

5.5.3 On the role of bottom up and top down saliency

Since reading chest x-rays is highly task dependent, it is generally believed that the bottom up saliency might not play an important role in attracting the fixations of the observers. On the contrary, the experimental results suggest that that the bottom up saliency plays a very significant and important role in attracting the fixations of the observers. This might be because pneumoconiosis is not a localized lung disease and the observers have to look at different lung regions to assess the profusion level in all the 6 zones. Unlike searching for localized lung tumor, profusion assessment involves more of a *scanning* strategy, instead of some *searching* strategy. The important role of bottom up saliency might be due to this scanning over different regions of chest x-ray. This might not be the case while searching for localized lung tumors.

Another reason for the important role of bottom up saliency might be that the anomalies related to pneumoconiosis also have higher intensity values compared to that of surrounding regions. Since our bottom up saliency model (GBVS) also uses higher intensity and orientation values to find the salient regions, GVBS model also predicts higher saliency over anomalies.

The temporal trends suggest that the role of bottom up saliency is stronger during the initial fixations than during the latter fixations, except in the case of staff radiologists. This result along with the results on the inter and intra observer fixation consistency suggests that during the first few fixations, the observers are looking more at the bottom up salient regions, after which, different visual strategies are evolved based on the information gained during the first few fixations.

Since the task given to all the observers is the same, top down influence on visual fixations should be same across the observers during the first few fixations. Since both bottom up and top down influences (for different observers) are more similar during the initial few fixations, we expect higher inter-observer fixation consistency, during the first few fixations. Our results also indicated the same phenomenon. The divergence of visual strategies after the initial few fixations might be due to either a reduced role of bottom up saliency and or the divergent top down visual strategies resulting from the information gained during the first few fixations. Since the role of bottom up saliency is not changing as much as the inter-observer fixation consistency, we can say that the top down visual strategies seems to be playing an important role in guiding the fixations of the observers after the initial few fixations. The relative role of top down and bottom up influence of visual fixations is still not completely understood and it is a difficult problem to address.

Several studies have proposed different psychological models to explain the relative importance of top down and bottom up visual saliency [83]. Tatler et. al. [83] proposed strategic divergence model of visual attention while free viewing the natural images. According to this model, there is no change in the role of bottom up saliency during the free viewing of an image, whereas observers use different top down strategies during the course of viewing. In the present study, even though different observers seem to use different top down strategies during the viewing of an image, the role of bottom up saliency is also changing with the number of fixations. Further experimental investigations are required to determine the relative importance of top down and bottom up visual saliency, while reading the chest x-rays of pneumoconiosis.

5.6 Concluding Remarks

The aim of the present study was to get insights into the factors that attract or guide the visual fixations of the observers while reading chest x-rays of pneumoconiosis. The reasonably good inter and intra observer fixation consistency suggests the use of similar viewing strategies. The viewing strategies of higher expertize groups seems to be more diverse compared to that of lower expertize groups. Doctors' visual strategies seem to vary with the image content, whereas the non-doctors' strategies vary far less with the image content. Thus, experience is helping the observers to develop new visual strategies based on the image content so that they can quickly and efficiently assess the disease level. First few fixations seem to be playing an important role in choosing the visual strategy, appropriate for the given image.

As expected, pneumoconiosis being a diffused lung disease, lung regions attract most of the gaze of the observers, whereas both rib and inter-rib regions are fixated almost equally. Thus, even though ribs carry less information about the disease, they are given equal importance as that of inter-rib regions.

The bottom up saliency (GBVS [32]) is shown to play an important role in attracting the fixations of the observers. This role of bottom up saliency seems to be more in lower expertize groups compared to that of higher expertize groups. Both bottom up and top down influence of visual fixations were found to change with time. Whereas, the role of bottom up influence is more during the initial few fixations, the role of top down influence seems to be more during the latter part of the viewing. The relative role of top down and bottom up influences of visual attention is still not completely understood and it remains the part of future work.

Chapter 6

Towards a new saliency model

Developing an automated gaze prediction system is useful for the development of CAD systems and training tools for the resident radiologists. Directing the observers' gaze towards salient regions of an x-ray can aid in diagnosis. Based on our experimental results, we have developed a new saliency model by combining the bottom up saliency and the saliency of lung regions in a chest x-ray. This new saliency model performed significantly better than bottom-up saliency in predicting the gaze of the observers in our experiment. This chapter discusses the details of this new saliency model. In section-6.1, we discuss about our new model of saliency, which we called Anatomical graph based visual saliency. This will be followed by some discussion and conclusion in sections 6.2 and 6.3 respectively.

6.1 Extended graph based visual saliency

In the previous chapter, it has been shown that bottom up saliency as predicted by Graph based visual saliency (GBVS) [32] model plays a significant role in attracting the gaze of the observers. It was also shown that lung regions attract most of the attention, which is generally expected with Pneumoconiosis being a diffused lung disease. This influence of lung regions in attracting the gaze of the observers can be considered as a top-down influence, as the importance of lung regions is mainly based on the task of the observers rather than the underlying image features. We also found that both inter-rib and rib regions are given equal importance by the observers.

Can we get better prediction of the observer saliency by combining the bottom-up saliency as predicted by GBVS model and the top-down saliency which is the importance of lung regions? We tried to answer this question, by modifying the GBVS saliency maps with the importance given to lung regions. We call these new modified maps as *Extended Graph Based Visual Saliency* (**EGBVS**) maps.

Figure 6.1 shows different steps in computing EGBVS saliency map from a sample chest x-ray image. First, a GBVS saliency map is computed from the given chest x-ray. Then, the lung regions are segmented from the chest x-ray, using the procedure explained in section-5.3.1, and a segmented lung mask is created as shown in figure 6.1.

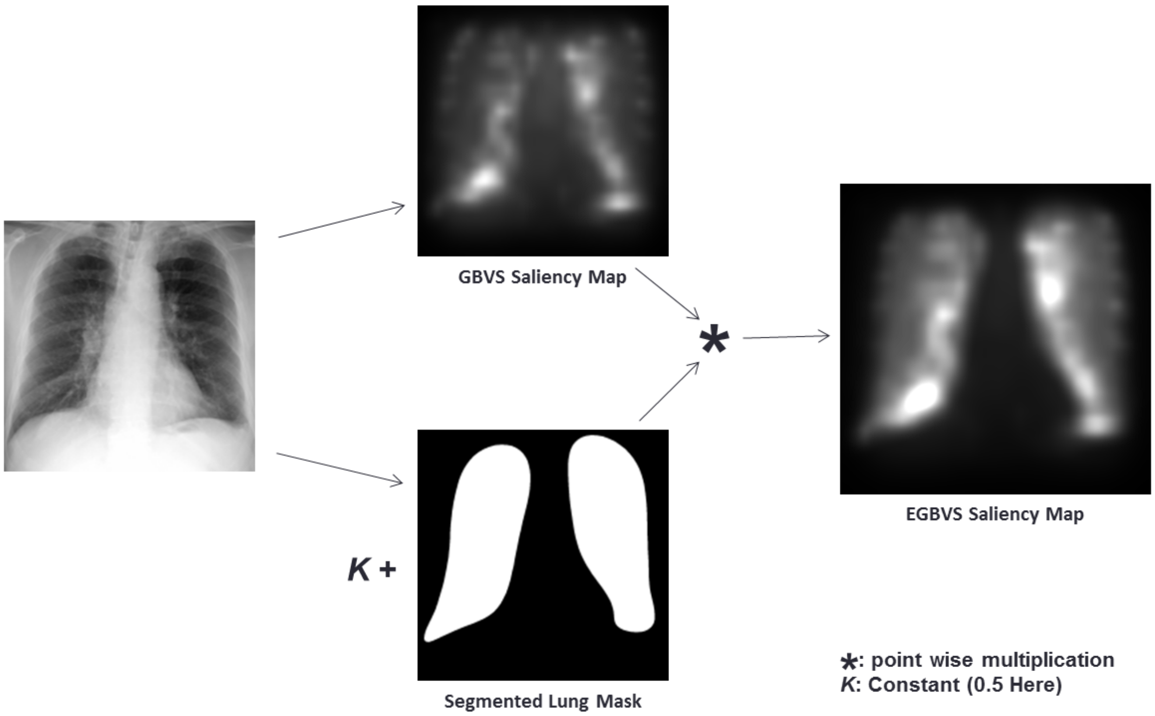


Figure 6.1 Different steps in extracting EGBVS saliency map from a sample chest x-ray

The next important step is regarding modifying the GBVS saliency map with the segmented lung mask. We have to modify GBVS saliency maps in such a way that the saliency inside lung regions should be increased whereas saliency outside lung regions should be decreased. That is, the pixels values of GBVS saliency map should be increased inside lung regions and the pixel values should be decreased in non-lung regions. We take our cue from a similar attempt to combine top-down and bottom-up influences in [73]. There, the task was to predict the eye fixations of the observers in interactive tasks like playing video games. The authors in [73] modeled bottom-up and top-down influences on eye position using separate saliency maps and then used a simple point-wise multiplication to combine bottom-up and top-down saliency maps. We also used same approach to modify GBVS saliency maps with segmented lung masks. The final EGBVS saliency map is obtained by combining GBVS saliency map and lung mask as follows:

$$\text{EGBVS Saliency Map} = (\text{GBVS Saliency map}) .* (\text{Segmented lung mask} + K)$$

Where $.*$ represents point wise multiplication and K represents a positive real constant between 0 and 1. In segmented lung mask images, lung regions are represented by pixel values of ‘1’ (white region) and remaining regions are represented by ‘0’ (black region). The K value is added to each of the pixels in lung segmented mask and is multiplied point-wise with GBVS saliency map to get EGBVS saliency map. It is clear that the value of K determines the relative importance given to lung and non-lung regions in final EGBVS saliency map. For example, if $K = 0.25$, all the GBVS saliency map pixels inside lung regions will be multiplied by 1.25, whereas all the pixel values in non-lung regions will be multiplied by 0.25. Any K value less than 1 would serve our purpose of decreasing saliency in non-lung regions and increasing saliency in lung regions. For K values greater than 1, saliency is not decreased in non-lung regions. Figure 6.2 shows EGBVS saliency maps obtained from a sample chest x-ray, when different K values are used. For $K = 0$, non-lung regions are segmented out of GBVS saliency maps. The saliency in non-lung regions increase as the value of K increases.

We experimented with different values of K and we found that values around 0.5 gave good prediction accuracy for final saliency maps. So, final EGBVS saliency maps are computed with K value as 0.5. Figure 6.3 shows some sample chest x-rays and the corresponding EGBVS saliency maps computed with $K = 0.5$.

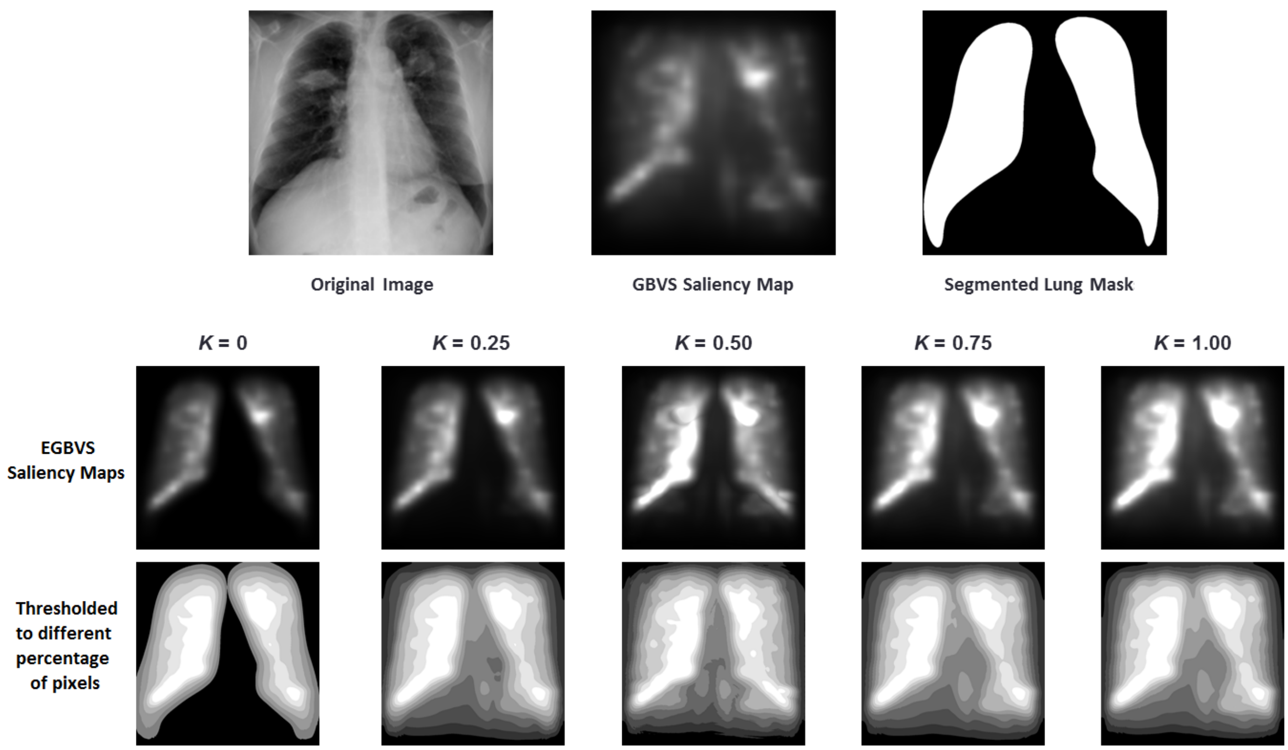


Figure 6.2 A sample chest x-ray and its EGBVS saliency maps for different values of K

6.1.1 Assessment of Proposed Model

In order to quantitatively measure the accuracy of EGBVS saliency maps in predicting the observers eye fixations, we used the same ROC analysis as explained in the previous chapter. By considering the saliency maps as a classifier on the fixation data AUC values are computed. These AUCs (averaged over

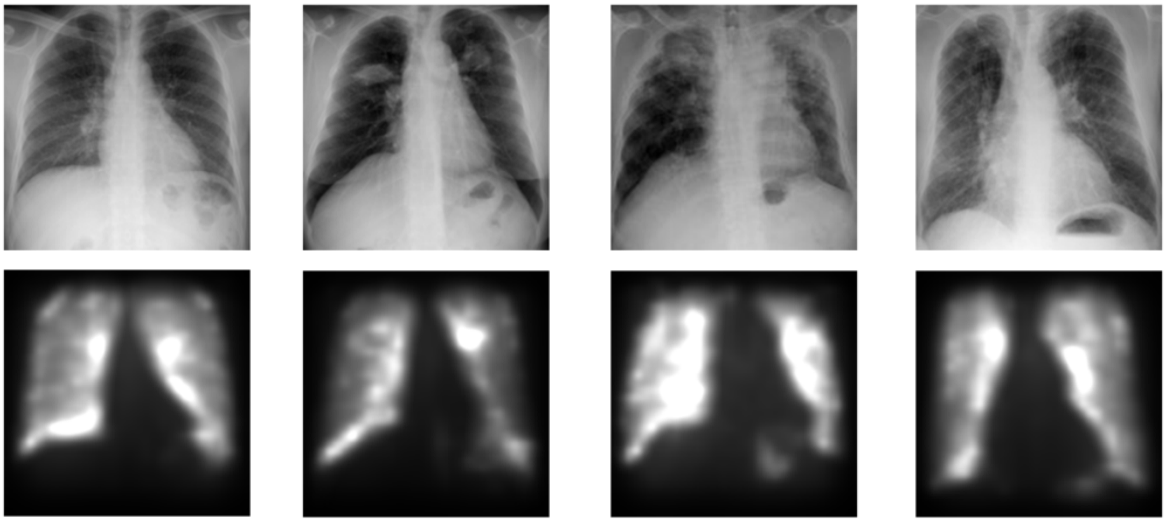


Figure 6.3 Sample chest x-rays (above) and corresponding EGBVS saliency maps (below)

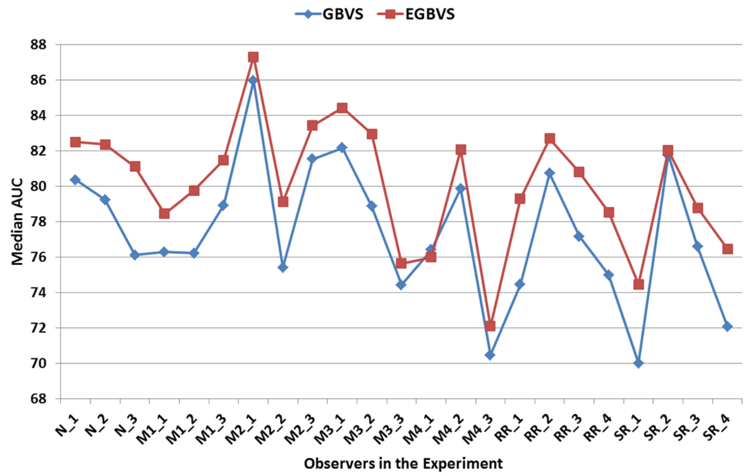
the image data set) indicate the performance of the saliency model in predicting the eye fixations of the observers.

Figure 6.4 shows the median AUCs, corresponding to both GBVS and EGBVS saliency maps, for all the participants in the experiment and for different expertize groups. From the above chart in figure 6.4, we can clearly see higher AUCs for EGBVS saliency maps than those of GBVS saliency maps, for all the participants. These charts also clearly show the performance difference between GBVS and EGBVS models, for different expertize groups. Wilcoxon signed rank test showed that the AUCs for EGBVS ($Mdn = 81.3$) are significantly higher ($Z = 2.0, p < 0.001$) than those of GBVS ($Mdn = 77.1$), for all the observers. In the previous chapter, we have already seen that GBVS model predicts the observers fixations to a good accuracy. EGBVS saliency model performs even significantly better (5.4% increase in median AUC) than GBVS model in predicting the eye fixations of the observers.

6.1.2 Comparisons with inter and intra-observer fixation consistency

Figure 6.5 shows the AUCs for EGBVS saliency maps along with inter-observer, intra-observer and random lung saliency maps, for different expertize groups. This figure also clearly shows the performance difference between GBVS and EGBVS models, for different expertize gropus. When considering all the observers, the Wilcoxon signed rank test showed that the AUCs for EGBVS model ($Mdn = 81.3$) are significantly higher ($Z = 18.5, p < 0.001$) than those related to inter-observer fixation consistency ($Mdn = 79$). There is no significant difference ($p = 0.67$) between the AUCs of EGBVS model and those of intra-observer consistency ($Mdn = 80.1$). There is no significant difference ($p = 0.410$) of EGBVS AUCs across different expertize groups.

Thus, EGBVS saliency maps perform significantly better (2.9% increase in median AUC) than the human saliency maps (of other observers) in predicting the eye fixations of the observers. Even though



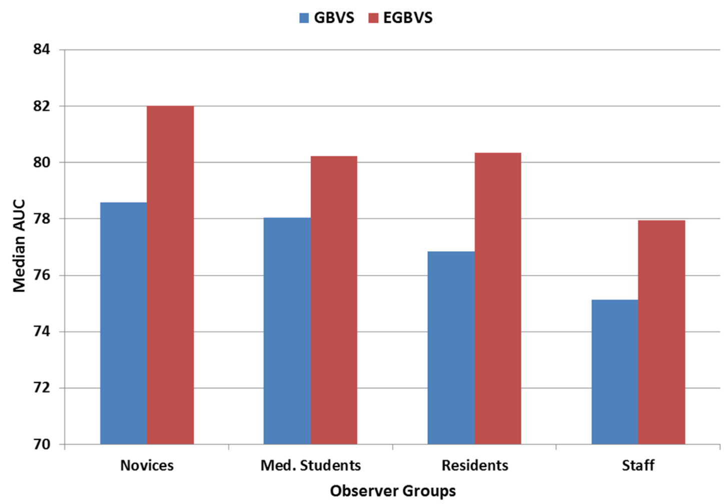


Figure 6.4 Median AUCs corresponding to GBVS and EGBVS saliency maps, (above) for all the participants in the experiment and (below) for different expertize groups.

there is no significant difference of AUCs for EGBVS saliency maps across different expertize groups, their role seems to be more in lower expertize groups compared to that of higher expertize groups.

6.2 Discussion

From the above results, it is clear that modifying the bottom-up saliency with more importance to lung regions can improve the fixation prediction accuracy. In fact, comparisons with AUCs related to intra and inter observer fixation consistency revealed that the bottom-up saliency and the importance to lung regions can explain most of the fixation consistency across the observers.

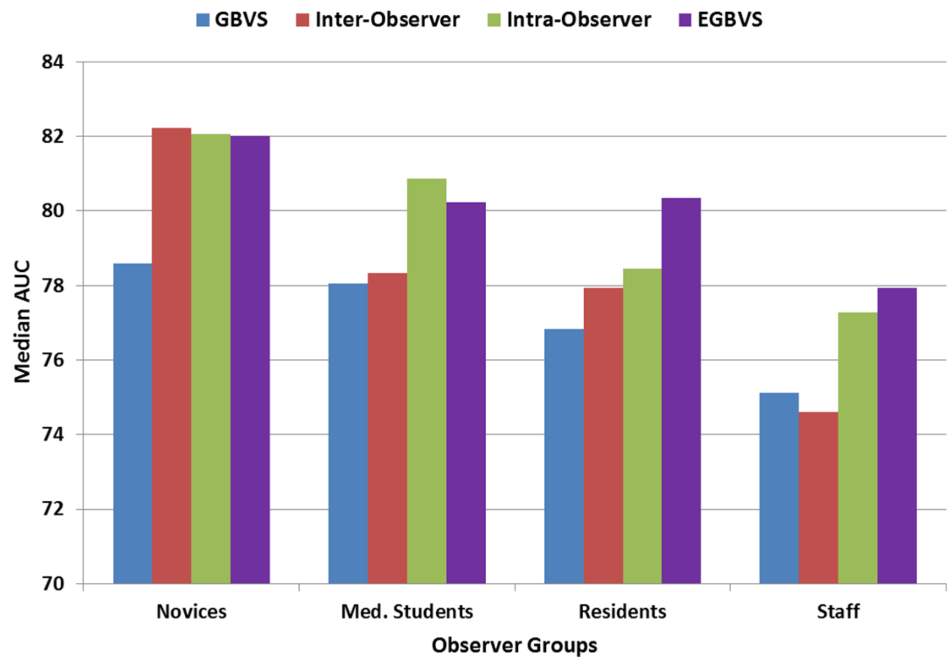


Figure 6.5 Median ROC areas, for different expertize groups, corresponding to Inter-observer & Intra-observer fixation consistency; and also those corresponding to GBVS and EGBVS saliency maps

We considered GBVS as our bottom-up model and segmented lung masks as top-down influence of attention. The lung masks capture the task given to the user, which is to assess the profusion rating of different zones *inside* lung regions. Since, in the previous chapter, we found that rib and inter-rib regions are fixated almost equally, these were not given special importance while considering the top-down influences. We used a simple point-wise multiplication to combine the bottom-up and top-down influences. Further experimentation is required to see if other strategies for combining the two types of information would lead to better fixation prediction accuracy.

There are many other potential top-down influences of attention, which can be considered to further improve the prediction accuracy, such as the role of contralateral symmetry and the relative importance of the different lung zones etc. The present model is just a starting point towards this direction of incorporating top-down influences to improve the fixation prediction accuracy of bottom-up models. We tried to quantitatively measure the contralateral symmetry (CS) in an x-ray image, and used that information in further modifying the bottom-up saliency maps. But, these were unsuccessful. This might be because the chosen quantitative measurements of contralateral symmetry did not reflect the CS information used by the observers. Alternatively, it could be that CS information might not have an important role to play in attracting the gaze of the observers. Further experiments are required to determine the role of CS information in guiding the fixations of the observers.

6.3 Concluding Remarks

In this chapter, we have introduced a new EGBVS saliency model which can predict the fixations of the observers to a good accuracy. Even though the model is a simple combination of bottom-up saliency maps and segmented lung masks, this demonstrates that even basic models using simple image features can predict the fixations of the observers to a good accuracy.

The main aim of this study is to show that the fixation prediction accuracy can be improved significantly by incorporating top-down influences on bottom-up saliency maps. We believe that further improvements can be made to this model by using other top-down influences such as the role of CS information, influences specific to expertize etc. These remain the part of future work.

Chapter 7

Conclusion and future directions

Pneumoconiosis, a common and serious occupational lung disease in developing countries, is mainly diagnosed through chest x-rays. The assessment of pneumoconiosis from chest x-rays is a complex process and it requires a good level of expertise. Most of the existing perceptual research on chest x-rays is on localized lung diseases such as lung cancer. Given that Pneumoconiosis is a diffused lung disease, many of these results might not be valid for chest x-rays of pneumoconiosis. In the present perception study, we studied the role of some important factors on various aspects of observers behavior such as diagnostic error, time and eye movements of the observers. We also did some in-depth analysis of eye movements to find the inter and intra observer consistency, the roles of top-down & bottom-up influences and, to get some insights into the visual strategies of various expertize groups.

First, we analyzed the role of expertize and contralateral symmetry (CS) on the diagnostic performance and eye movements of the observers. Results indicate that expertise and CS play important roles in the diagnosis of pneumoconiosis. CS seems to be helping in reducing the general tendency of giving less profusion ratings. For residents, the eye scanning strategies seem to play an important role in using the CS information present in chest radiographs; however, in staff radiologists, peripheral vision or higher level cognitive processes seems to play a role in using the CS information.

Further experiments are required to determine the exact role of CS in reading chest x-rays i.e. how exactly this information is being used by the radiologists. We still do not know if this CS information is useful at the image level or zonal level or at intra-rib region level. Understanding this would help in incorporating CS based image features in developing better computer aided diagnostic tools.

Next, we did an in-depth analysis on the eye movements of the observers. We found good consistency of fixations with-in and across observers, showing the importance of information underlying the fixations. Lower expertize groups seem to be using same visual strategies independent of image content, whereas higher expertise groups are able to develop different visual strategies depending on the image content, so that they can quickly and efficiently assess the disease level. First few fixations seem to be playing an important role in choosing the visual strategy, appropriate for the given image.

Even though assessing pneumoconiosis is a specialized task, the bottom-up saliency seems to be playing a very important role in attracting the fixations of the observers. This is not an entirely surprising result, given the fact that the observers have to scan through all the zones of the chest x-ray, to assess the profusion level. This result might not be valid for localized lung diseases, where radiologists to have to localize lung tumors or nodules. This role of bottom up saliency seems to be more in lower expertize groups compared to that of higher expertize groups. Despite the popular belief that the inter-rib regions are generally given more importance by the radiologists, we found that both rib and inter-rib regions are fixated almost equally. Pneumoconiosis, being a lung disease, lung regions attracts most of the attention. Both bottom up and top down influence (importance of lung regions) of visual fixations were found to change with time. Whereas, the role of bottom up influence is more during the initial few fixations, the role of top down influence seems to be more during the latter part of the viewing. A good future work direction, in this regard, would be to study the other aspects of top-down influences such as the role of CS information, in guiding the fixations of the observers. Also, finding the relative roles of top-down and bottom-up influences on eye fixations is a part of future work.

Finally, based on the experiments results, we introduced a simple saliency model by modifying the bottom-up saliency maps with segmented lung masks. The present model is just a starting point towards the direction of incorporating top-down influences to improve the fixation prediction accuracy of bottom-up models. We believe that further improvements can be made to this model by using other top-down influences such as the role of CS information, influences specific to expertize etc. These remain the part of future work.

Although our study is specific to chest x-rays of Pneumoconiosis, our eye tracking experiments mainly involve the radiologists assessing the *profusion level* of different lung zones but not the final assessment of Pneumoconiosis. Final assessment of Pneumoconiosis for a given x-ray image is based on patient working history, profusion levels on different lung zones and size of the radio opacities. The observers were not asked to report the size of the opacities as profusion category is of primary importance in assessing the disease level. The observers were also ignorant of patients' working history. Thus, the present experimental study is more about assessing the profusion level of radio opacities rather than assessing Pneumoconiosis. Hence, many of the results obtained in the present study can be extended to other interstitial lung diseases which are diagnosed through chest x-rays and where abnormalities are of diffused nature like that of Pneumoconiosis.

Overall, we believe that we made some significant contributions in furthering our present understanding of some perceptual and cognitive factors involved in the diagnosis of pneumoconiosis. Some of the present findings can be used in developing better training regimes or developing better computer aided diagnostic tools. For example, the present study showed that the careful analysis of all the lung zones is required for good diagnostic results. We can also develop some computer aided tools with emphasis on bottom-up salient regions and using CS information and see if such tools can improve the diagnostic accuracy. Many of the results in the present work might not be valid for reading the chest x-rays of localized lung diseases such as cancer. It would be interesting to find out which of these experimental results are valid for diagnosing localized lung diseases as well.

Related Publications

- V. Jampani, Ujjwal and J. Sivaswamy, Assessment of Computational Visual Attention Models on Medical Images. *Indian Conference on Vision, Graphics and Image Processing*, Mumbai, India, Dec. 2012. (to be published)
- V. Jampani, V. Vaidya, J. Sivaswamy, and L.T. Kishore. Role of expertise and contralateral symmetry in the diagnosis of pneumoconiosis: an experimental study, *Proc. of SPIE 2011*, Vol. 7966, March 2011.
- V. Jampani, V. Vaidya, J. Sivaswamy, P. Ajemba, and L.T. Kishore, Effect of expertise and contralateral symmetry on the eye movements of observers while diagnosing pneumoconiosis, *Medical Image Perception Society Conference*, Dublin, August 2011 (MIPS Student Scholar).

Bibliography

[1] Adaptive Whitening Saliency Matlab Code. http://www-gva.dec.usc.es/persoal/xose.vidal/research/aws/AWSmodel.html.

[2] Context Aware Saliency Matlab Code. http://cgm.technion.ac.il/Computer-Graphics-Multimedia/Software/Saliency/Saliency.html.

[3] Dynamic Visual Attention Saliency Matlab Code. http://www.klab.caltech.edu/~xhou/projects/dva/dva.html.

[4] Graph Based Visual Saliency Matlab Code. http://www.klab.caltech.edu/~harel/share/gbvs.php.

[5] Image Signature Saliency Matlab Code. http://www.klab.caltech.edu/~harel/share/gbvs.php.

[6] Information Maximization Saliency Matlab Code. http://www.klab.caltech.edu/~harel/share/gbvs.php.

[7] Itti-Koch Saliency Model Matlab Code. www.saliencytoolbox.net.

[8] Self Resemblance Saliency Matlab Code. http://users.soe.ucsc.edu/~milanfar/research/rokaf/.html/download.php.

[9] Spectral Residual Saliency Matlab Code. http://www.klab.caltech.edu/~xhou/projects/spectralResidual/spectralresidual.html.

[10] SR Research: Eye Link 1000. http://www.sr-research.com/EL_1000.html.

[11] C. Anderson and D. Van Essen. Shifter circuits: a computational strategy for dynamic aspects of visual processing. *Proceedings of the National Academy of Sciences*, 84(17):6297–6301, 1987.

[12] P. Annangi, S. Thiruvenkadam, A. Raja, H. Xu, X. Sun, and L. Mao. A region based active contour method for x-ray lung segmentation using prior shape and low level features. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 892–895. IEEE, 2010.

[13] M. Antonelli and G. Yang. Lung nodule detection using eye-tracking. In *IEEE International Conference on Image Processing, ICIP 2007*, volume 2, pages 457–460. IEEE, 2007.

[14] G. Backer, B. Mertsching, and M. Bollmann. Data-and model-driven gaze control for an active-vision system. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(12):1415–1429, 2001.

[15] F. Baluch and L. Itti. Mechanisms of top-down attention. *Trends in neurosciences*, 34(4):210–224, 2011.

[16] I. Bankman. *Handbook of medical imaging: processing and analysis*. Academic Press, 2000.

[17] N. Bruce and J. Tsotsos. Saliency, attention, and visual search: An information theoretic approach. *Journal of Vision*, 9(3), 2009.

[18] C. Chen, K. Kao, and C. Tyler. Face configuration processing in the human brain: the role of symmetry. *Cerebral Cortex*, 17(6):1423–1432, 2007.

[19] E. Christensen, R. Murry, K. Holland, J. Reynolds, M. Landay, and J. Moore. The effect of search time on perception. *Radiology*, 138(2):361–365, 1981.

[20] A. Cochrane and L. Garland. Observer error in the interpretation of chest films; an international investigation. *Lancet*, 2(6733):505–509, 1952.

[21] I. R. P. A. I. N.-I. R. Committee, I. L. Organisation, and I. L. Office. *Guidelines for the Use of ILO International Classification of Radiographs of Pneumoconioses*. International Labour Office, 1980.

[22] J. de Valk and E. Eijkman. Analysis of eye fixations during the diagnostic interpretation of chest radiographs. *Medical and Biological Engineering and Computing*, 22(4):353–360, 1984.

[23] L. Dempere-Marco, X. Hu, S. MacDonald, S. Ellis, D. Hansell, and G. Yang. The use of visual search for knowledge gathering in image decision support. *Medical Imaging, IEEE Transactions on*, 21(7):741–754, 2002.

[24] A. Duchowski. *Eye tracking methodology: Theory and practice*. Springer-Verlag New York Inc, 2007.

[25] S. Frintrop. *VOCUS: A visual attention system for object detection and goal-directed search*, volume 3899. Springer-Verlag New York Inc, 2006.

[26] A. Garcia-Diaz, X. Fdez-Vidal, X. Pardo, and R. Dosil. Saliency based on decorrelation and distinctiveness of local responses. In *Computer Analysis of Images and Patterns*, pages 261–268. Springer, 2009.

[27] L. Garland. On the scientific evaluation of diagnostic procedures. *Radiology*, 52(3):309–328, 1949.

[28] L. Garland. On the reliability of roentgen survey procedures. *The American journal of roentgenology and radium therapy*, 64(1):32–41, 1950.

[29] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2376–2383. IEEE, 2010.

[30] K. Grammer and R. Thornhill. Human (homo sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. *Journal of Comparative Psychology; Journal of Comparative Psychology*, 108(3):233–242, 1994.

[31] D. Hansen and J. Qiang. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(3):478–500, 2010.

[32] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. *Proceedings in neural information processing systems*, 19, 2006.

[33] D. Heinke and G. Humphreys. Saim: A model of visual attention and neglect. *Artificial Neural NetworksI-CANN'97*, pages 913–918, 1997.

[34] D. Heinke and G. Humphreys. Computational models of visual selective attention: A review. *Connectionist models in cognitive psychology*, 1(Part 4):273–312, 2005.

[35] X. Hou, J. Harel, and C. Koch. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):194, 2012.

[36] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[37] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. *Advances in neural information processing systems*, 21:681–688, 2008.

[38] C. Hu, H. Kundel, C. Nodine, E. Krupinski, and L. Toto. Searching for bone fractures: a comparison with pulmonary nodule search. *Academic Radiology*, 1(1):25–32, 1994.

[39] G. Humphreys and H. Muller. Search via recursive rejection (serr): A connectionist model of visual search. *Cognitive Psychology*, 25(1):43–110, 1993.

[40] L. Itti and C. Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10-12):1489–1506, 2000.

[41] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2106–2113. IEEE, 2009.

[42] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human neurobiology*, 4(4):219–227, 1985.

[43] S. Kollmorgen, N. Nortmann, S. Schröder, and P. König. Influence of low-level stimulus features, task dependent factors, and spatial biases on overt visual attention. *PLoS computational biology*, 6(5):e1000791, 2010.

[44] R. Kruger, W. Thompson, and A. Turner. Computer diagnosis of pneumoconiosis. *Systems, Man and Cybernetics, IEEE Transactions on*, (1):40–49, 1974.

[45] E. Krupinski. The importance of perception research in medical imaging. *Radiation medicine*, 18(6):329–334, 2000.

[46] E. Krupinski, H. Kundel, P. Judy, and C. Nodine. Key issues for image perception research. *Radiology*, 209(3):611–612, 1998.

[47] H. Kundel. Images, image quality and observer performance. *Radiology*, 132(2):265–271, 1979.

[48] H. Kundel. Visual search in medical images. *In Handbook of Medical Imaging: Physics and Psychophysics*, 2000.

[49] H. Kundel. Reader error, object recognition, and visual search. In *Proceedings of SPIE*, volume 5372, pages 1–9, 2004.

[50] H. Kundel. History of research in medical image perception. *Journal of the American College of Radiology*, 3(6):402–408, 2006.

[51] H. Kundel and P. La Follette Jr. Visual search patterns and experience with radiological images. *Radiology*, 103(3):523–528, 1972.

[52] H. Kundel and C. Nodine. Interpreting chest radiographs without visual search. *Radiology*, 116(3):527–532, 1975.

[53] H. Kundel, C. Nodine, D. Carmody, et al. Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, 13(3):175–181, 1978.

[54] H. Kundel, C. Nodine, E. Conant, and S. Weinstein. Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, 242(2):396–402, 2007.

[55] H. Kundel and D. Wright. The influence of prior knowledge on visual search strategies during the viewing of chest radioqraphs. *Radiology*, 93(2):315–320, 1969.

[56] Q. Li, S. Katsuragawa, and K. Doi. Improved contralateral subtraction images by use of elastic matching technique. *Medical Physics*, 27:1934–1942, 2000.

[57] Q. Li, S. Katsuragawa, T. Ishida, H. Yoshida, S. Tsukuda, H. MacMahon, and K. Doi. Contralateral subtraction: A novel technique for detection of asymmetric abnormalities on digital chest radiographs. *Medical physics*, 27:47–55, 2000.

[58] F. Liddell. An experiment in film reading. *British Journal of Industrial Medicine*, 20(4):300–312, 1963.

[59] M. Loog and B. Ginneken. Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification. *Medical Imaging, IEEE Transactions on*, 25(5):602–611, 2006.

[60] D. Manning, A. Gale, and E. Krupinski. Perception research in medical imaging. *British journal of radiology*, 78(932):683–685, 2005.

[61] R. Mason and V. Broaddus. *Murray and Nadel's textbook of respiratory medicine*. Philadelphia: Elsevier Saunders, 2005.

[62] H. Matsumoto, Y. Terao, A. Yugeta, H. Fukuda, M. Emoto, T. Furubayashi, T. Okano, R. Hanajima, and Y. Ugawa. Where do neurologists look when viewing brain ct images? an eye-tracking study involving stroke cases. *PloS one*, 6(12):e28928, 2011.

[63] R. Milanese. *Detecting salient regions in an image: from biological evidence to computer implementation*. PhD thesis, University of Geneva, 1993.

[64] R. Milanese, H. Wechsler, S. Gill, J. Bost, and T. Pun. Integration of bottom-up and top-down cues for visual attention using non-linear relaxation. In *In Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 781–785. IEEE, 1994.

[65] R. Morgan, M. Donner, B. Gayler, S. Margulies, P. Rao, and P. Wheeler. Decision processes and observer error in the diagnosis of pneumoconiosis by chest roentgenography. *American Journal of Roentgenology*, 117(4):757–764, 1973.

[66] M. Myles-Worsley, W. Johnston, and M. Simons. The influence of expertise on x-ray image processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(3):553–557, 1988.

[67] V. Navalpakkam and L. Itti. A goal oriented attention guidance model. In *Biologically Motivated Computer Vision*, pages 81–118. Springer, 2002.

[68] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision research*, 45(2):205–231, 2005.

[69] C. Nodine and H. Kundel. The cognitive side of visual search in radiology. *Eye Movements: From Psychology to Cognition. North Holland, Elsevier Science*, pages 572–582, 1987.

[70] C. Nodine, H. Kundel, et al. Using eye movements to study visual search and to improve tumor detection. *Radiographics*, 7(6):1241–1250, 1987.

[71] J. Oestmann, R. Greene, D. Kushner, P. Bourgouin, L. Linetsky, and H. Llewellyn. Lung lesions: correlation between viewing time and detection. *Radiology*, 166(2):451–453, 1988.

[72] N. Pearson, J. Ashford, D. Morgan, R. Pasqual, and S. Rae. Effect of quality of chest radiographs on the categorization of coalworkers' pneumoconiosis. *British Journal of Industrial Medicine*, 22(2):81–92, 1965.

[73] R. Peters and L. Itti. Beyond bottom-up: Incorporating task-dependent influences into a computational model of spatial attention. In *Computer Vision and Pattern Recognition*, pages 1–8, 2007.

[74] R. Phaf, A. Van der Heijden, and P. Hudson. Slam: A connectionist model for attention in visual selection tasks. *Cognitive Psychology*, 22(3):273–341, 1990.

[75] R. Reger, C. Smith, J. Kibelstis, and W. Morgan. The effect of film quality and other factors on the roentgenographic categorization of coal workers' pneumoconiosis. *The American journal of roentgenology, radium therapy, and nuclear medicine*, 115(3):462–472, 1972.

[76] J. Reynolds and D. Heeger. The normalization model of attention. *Neuron*, 61(2):168–185, 2009.

[77] S. Rockoff and A. Schwartz. Roentgenographic underestimation of early asbestosis by international labor organization classification. analysis of data and probabilities. *Chest*, 93(5):1088–1091, 1988.

[78] F. Sani, J. Todman, and J. Todman. *Experimental design and statistics for psychology: a first course*. Wiley-Blackwell, 2006.

[79] H. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*, 9(12):1–27.

[80] F. Shic and B. Scassellati. A behavioral analysis of computational models of visual attention. *International Journal of Computer Vision*, 73(2):159–177, 2007.

[81] K. Subba Rao. Personal communication. 2009.

[82] M. Tacca. Commonalities between perception and cognition. *Frontiers in Psychology*, 2, 2011.

[83] B. Tatler, R. Baddeley, and I. Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision research*, 45(5):643–659, 2005.

[84] E. Thomas and E. Lansdown. Visual search patterns of radiologists in training. *Radiology*, 81(2):288–292, 1963.

[85] A. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1):97–136, 1980.

[86] J. Tsotsos, S. Culhane, W. Kei Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1):507–545, 1995.

[87] S. Tsukuda, A. Heshiki, S. Katsuragawa, Q. Li, H. MacMahon, and K. Doi. Detection of lung nodules on digital chest radiographs: Potential usefulness of a new contralateral subtraction technique. *Radiology*, 223(1):199–203, 2002.

[88] B. Van Ginneken, B. ter Haar Romeny, and M. Viergever. Computer-aided diagnosis in chest radiography: a survey. *Medical Imaging, IEEE Transactions on*, 20(12):1228–1241, 2001.

[89] X. Wang and D. Christiani. Occupational lung disease in china. *International journal of occupational and environmental health*, 9(4):320–325, 2003.

[90] J. Wolfe. Visual attention. *Seeing*, 2:335–386, 2000.

[91] J. Wolfe, K. Cave, and S. Franzel. Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology-human Perception and Performance*, 15:419–433, 1989.

[92] L. Wong and H. Ewe. A study of lung cancer detection using chest x-ray images. In *Proceedings of 3rd APT Telemedicine Workshop, Kuala Lumpur*, pages 27–28, 2005.

[93] R. Wright and L. Ward. *Orienting of attention*. Oxford University Press, USA, 2008.

[94] A. Yarbus. *Eye movements and vision*. Plenum press, 1967.

[95] H. Yoshida and K. Doi. Computerized detection of pulmonary nodules in chest radiographs: reduction of false positives based on symmetry between left and right lungs. In *Proceedings of SPIE*, volume 3979, page 97, 2000.