THE UNIVERSITY OF NOTTINGHAM

School of Electrical and Electronic Engineering

NEURAL NETWORK EDGE DETECTION AND SKIN LESIONS IMAGE SEGMENTATION METHODS: ANALYSIS AND EVALUATION

By

Maher I. Rajab

Thesis Submitted to the University of Nottingham
For the Degree of Doctor of Philosophy

September 2003

Table of Contents

Acknowledgements

First of all, I would like to express my deepest gratitude to my supervisor Dr Malcolm S. Woolfson. He patiently guided me through the course of my PhD study program. His gracious encouragement, and enthusiasm during the course of this PhD has made the completion of this thesis visible. I would also like to extend my grateful thanks to Dr Steve P. Morgan for sharing his biomedical knowledge and scientific research methodologies. He was instrumental in completing the journal publication.

Special thanks are reserved to Dr K. M. Curtis and to my friend Dr H. Amin (Bradford Univ.). My appreciation goes to Dr Chung See for his great comments and advice through my yearly technical reports. To B. van der Zwaag, Systems and Signals, Univ. of Twente, Netherlands, for his e-mail contacts during some of the verification stage of my work. Thanks to Mr Raymond J. Brooks for his assistance in English proofing and his continued patience.

I express my regards and thanks to my parents for their continual encouragement and to my wife for her patience and care during my stay in Nottingham. To my little son Ammar who gave positive meaning to our life with his lively and bright eyes.

Finally, I would like to thank Astron Clinica, Cambridge, UK, for the support in providing the real colour skin lesion images, which facilitate the work throughout this study. I would like to express my sincere thanks to the government of *Saudi Arabia* for the aid in sponsoring my PhD studentship.

Abstract

Similar to a human observer, an automated image vision system is able to recognise most parts of an object if the system could accurately trace and reflect its true shape. This has prompted the development of the many diverse edge detection techniques. Neural networks have been successfully applied to pattern recognition tasks and edge detection. However, there is a great necessity to analyse neural network models so as to achieve close insight into their internal functionality. To this purpose, a new and general training set, consisting of a limited number of prototype edge patterns, is proposed to analyse the problem of neural network edge detection.

This thesis also proposes two approaches to the skin lesion image segmentation problem. The first is a mainly thresholding segmentation method where an optimal threshold is determined iteratively by an isodata algorithm. The second method proposed is based on neural network edge detection and a rational Gaussian curve that fits an approximate closed elastic curve between the recognized neural network edge patterns. A quantitative comparison of the techniques is enabled by the use of synthetic lesions to which Gaussian noise is added. The proposed techniques are also compared with an established automatic skin segmentation method. It is demonstrated that for lesions with a range of different border irregularity properties the thresholding segmentation method provides the best performance over a range of signal to noise ratios; the thresholding segmentation method is also demonstrated to have similar performance when tested on real skin lesions.

CHAPTER 1: INTRODUCTION

The aim of this chapter is to state the motivation of the thesis, which is the accurate automatic delineation of melanoma lesion border shapes, which would improve the effectiveness of a computer-aided system for the analysis and classification of benign or malignant lesions. Section 1.1 introduces general issues concerning image segmentation and to provide a brief overview of the need for and the description of the work presented in this thesis. The vast field of image processing, along with the issues concerning image segmentation, are also presented. The particular problem of the segmentation of melanoma images is described. This will be followed by an outline of the objectives of the research and the tasks of the work, section 1.2. Section 1.3 includes the introduction of the new approaches presented within the thesis. Finally, an outline is presented of the remainder of the thesis contents.

1.1 Image processing

Digital image processing deals with the processing of digital images, which consists of a matrix of pixels representing intensities of various positions. The familiar example of these digitised images are those produced by a scanner or by a video camera. Once the image has been digitised, it can be operated upon by various image processing operations. The goals of these processing steps are the improvement of images for human understanding and the processing of scene data for automatic machine perception [Gokmen, 1990].

Image processing operations can be roughly divided into three major categories: Image Compression, Image Enhancement and Restoration, and Feature Extraction. Image compression is familiar to many engineers. It involves reducing the amount of memory needed to store a digital image. An application of image compression such as JPEG coding [Motta, May 2002] , is an advantage for transmitting files across networks and for archiving libraries of images. Image restoration deals with the problem of image defects, which could be caused by the digitisation process. Faults in the imaging set-up (for example, bad lighting) can be corrected using Image Enhancement techniques [Shyu and Leou, 1998]. Once the image is in good condition, the Feature Extraction operations can be used to obtain useful information from the image [Nixon and Aguado, 2002].

1.1.1 Image segmentation

The autonomous machine perception task is achieved by a number of steps. The initial step is the segmentation of the image into a meaningful region or object. When analysing a region or object in an image it is vital that we distinguish between the object of interest and the background. From this division, the object can be identified by its shape or from other features. The segmentation task usually starts with the extraction of the limits of the object. These limits are commonly called "edges". Moreover, an object contour, which may be constructed from the compacted information provided by these edges, can facilitate the measurements on the object.

1.1.2 Segmentation of medical images

Current computing research in medical image analysis uses image segmentation as an intermediate step to isolate the object of interest from the background [Olabarriaga and Smeulders, 2001]. In the area of biomedical image segmentation, most of the proposed techniques can be categorized into three classes [Fu and Mui, 1981], (1) characteristic feature thresholding or clustering, (2) edge detection, and (3) region extraction. Tasks such as measurements [Giachetti, et al., 2003], visualisation [Campbell and Flynn, 2001] and registration [Maintz and Viergever, 1998] are supported by segmentation. The need for an efficient segmentation is of great importance because of its potential need in the identification of important clinical morphological information about the patient's anatomy e.g. [Imeliska, et al., 2000] and pathology e.g. [Brenner, et al., 1981].

1.1.3 Segmentation of Melanoma Images

Malignant melanoma skin cancer is a potentially fatal disease, killing a disproportionate percentage of economically active younger people when compared with other human cancers. The highest incidence of skin cancer in the world is in Australia; Australians are eight times more likely to develop common skin cancer than any other form of cancer. A recent report has indicated that one out of every two Australians will develop skin cancer at some stage during their lives and over 720,000 skin cancer removal operations take place annually, at a cost to the Federal Government of over $300 million [Van der Hoek, 2000]. There is a steady increase in the incidence of this condition in many parts of the world [Van der Hoek, 2000]

including the USA [Jemal, et al., 2001] and Scandinavia [Cohn-Cedermark, et al., 2000]. Early detection of melanoma is feasible because it is usually visible on the skin surface when in a curable stage [Weinstock, 2000]. On the contrary, an invasive malignant melanoma (MM) reduces drastically the chance for the patient to survive. For instance, an epidemiological study for malignant melanoma in Scotland reported that survival fell in men from 93% for tumours thinner than 1.5 mm to 47% for tumours thicker than 3.5 mm, and in women from 97% to 55% respectively [MacKie, et al., 2002].

Sometimes physicians cannot visually distinguish between benign lesions and early MM. Therefore, a biopsy is performed and the pathologist diagnoses the lesion. However, several drawbacks of biopsies are encountered [Marquez, 2001]: (i) significant return-time for a biopsy, (ii) it can be painful for a patient to have several questionable lesions excised which, in some cases can be disfiguring, and (iii) incomplete biopsy might cause a spread or metastasise of melanoma cells to other parts of the body and these may invade other organs. The well recognized and established technique of dermatoscopy or epiluminescence microscopy (ELM) [Carli, et al., 2000;Goldman, 1951;Pehamberger, et al., 1993;Steiner, et al., 1987] is a noninvasive, in vivo technique which would eliminate the need for unnecessary or incorrect biopsies because, with ELM, dermatologists can view and image the subsurface layers of the skin. This technique reveals morphological features, colours and patterns not perceptible by the dermatologist during the clinical observation. For example, ELM permits in vivo visualization of pigmented anatomical structures of

the epidermis and, even beyond that, of the dermo-epidermal junction and superficial papillary dermis, which are impossible to be examined with the naked eye [Kent, et al., 1993]. Trained dermatologists in the use of ELM can improve their diagnostic accuracy of melanoma from about 65% using the unaided eye to approximately 80% with the benefit of ELM [Kanzler and Mraz-Gernhard, 2001]. However, even with ELM, a trained dermatologist can be deceived at least 20% of the time by the appearance of a melanoma [Kanzler and Mraz-Gernhard, 2001].

Low rate of correct classification of clinical diagnosis [Fisher and Schmid, 1996] calls for the development of both digitised ELM (DELM) and automated image analysis systems. Computer aided diagnosis systems have been successfully applied for early detection of melanoma. For example, a recently developed PC-based pilot system by Binder et al. [Binder, et al., 2000]. The system promises to automatically segment the digitised ELM images, measuring 107 morphological parameters. A neural network classifier trained with these features is able to differentiate between benign and malignant melanoma.

Accurate delineation of the lesion border shape may improve the effectiveness of a computer-aided system, similar to the above, and make them accessible for further analysis and classification (e.g. benign or malignant). Measurement of image features for diagnosis of malignant melanoma (a type of cancer of the skin) requires the first step to be the detection and localisation of the lesion in an image. The accurate determination of lesion boundaries is a primary aim in this work so that

measurement, e.g. asymmetry, border irregularity, colour variation, and diameters can be accurately computed [Xu, et al., 1999].

1.2 Research objectives

The initial objective is to investigate and compare the feasibility of two novel approaches for generalized edge detection in images. An analysis of the accurate delineation of border shapes in synthetic environments using several diverse methods has not been reported before. The main aim of this research is to develop a novel method of image segmentation and to compare it to the other methods in the literature. The development will be carried out in successive steps and is summarized as follows:

1- Starting with the preliminary experiments on synthetic environments to verify the edge detection algorithm that should be used prior to establishing the design of the proposed delineation algorithm.

2- A novel neural network training set is designed and tested on a variety of real images and with different training set sizes. Since the variations of intensities between image pixels is mostly dependent on the nature of the image itself, a careful selection of the training set being applied to the NN in the learning stage is required so that its output can be interpreted more correctly. The most successful training sets are chosen as the basis in the design of another approach to delineate skin lesions; neural network edge detection (NNED).

3- It may be hypothesized that the edges produce by different algorithms do not establish the correct maps. Thus it may be possible to override some of the vagueness in detecting the edges by comparison between these maps. **The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be quantitatively and more accurately compared.**

4- to test the ability of the derived methods to delineate the lesion border shape for real lesion images; the derived technique may improve the effectiveness of a computer-aided system for the analysis and classification of benign or malignant lesions. It should be noted that the variability between expert clinical delineations is dependent on human subjectivity (this will be discussed in Chapter 5). This variability was not addressed in many evaluation studies of automated skin lesions segmentation methods. Moreover, in spite of these variations between clinical delineations, they were always considered as the gold standard to evaluate skin segmentation methods (See Chapter 5). Therefore, our main objective of using real images is to aid the final stage of testing the derived methods rather than using any human contribution in any stages of the design or development of the derived segmentation methods.

1.3 Proposed solutions: NNED and IS

In this thesis we investigate the application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection

(NNED). The aim is to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al. is also used here to verify the other two proposed methods. These approaches are compared for synthetic lesions at different image signal to noise ratios (SNRs). *The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be quantitatively and more accurately compared.*

Preliminary experiments are carried out with a neural network edge detector applied to general real images after being trained with simple and small data sets, since the strategy of minimizing the complexity and ambiguity in the training set is a significant factor in the success of neural network recognition. The most successful training sets are chosen for the next stage of experiments on noisy synthetic lesion images as an attempt to establish the basis for the general neural network edge detection method (NNED).

In this work, the IS technique has been developed to extract the true border that reveals the global structure irregularity (indentations and protrusions), which may suggest excessive cell growth or regression of a melanoma. The algorithm is applied to the Blue channel of the RGB colour vectors to distinguish lesions from the skin and proceed with grey scale morphology and background noise reduction to enhance and filter the image of lesion. The algorithm also does not depend on the use of rigid

threshold values, because the isodata algorithm (section 3.3.2, Chapter 3) that is used determines an optimal threshold iteratively. Verification experiments are performed on digitised clinical photographs and also pigmented networks captured with the epiluminescence microscopy (ELM) technique. We demonstrate that we can also delineate pigmented networks in skin lesions, and make them accessible for further analysis and classification.

1.4 Outline of the thesis

This thesis consists of six chapters including this chapter. The contents of the remaining five chapters are described briefly as follows:

Chapter 2: In this chapter a literature review on topics relating to the thesis is presented. The areas that are directly related to this thesis will be covered in detail. NNs are investigated as one of the key elements used in this research. There are numerous kinds of NNs and also a wide range of their applications. Discussion of the applications of NNs to edge detection is particularly presented. The most widely used NNs architecture is chosen in this work and its learning algorithm, back propagation (BP) training algorithm. Finally, the mathematical background of another edge detection technique, optimal thresholding, is described which is the basic building block of the developed iterative segmentation (IS) algorithm.

Chapter 3: The three techniques that are applied in this research to the segmentation of pigmented skin lesions are described. We investigated the application of two

approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED). The aim is to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al. [Xu, et al., 1999] is also used here to verify the other two proposed methods.

Chapter 4: This chapter presents the practical implementation of the neural network edge detector (NNED), utilizing the analysis of the neural networks edge detector applied to synthetic noisy lesions. Extensive investigations of simple real images are also presented to test and validate the novel methods. A limited number of noisy edge patterns, is proposed to analyse the capability of neural networks edge detection for both synthetic images and real images. Various methods are applied to reduce the neural network noisy training set; e.g. variation of training set size.

Chapter 5: The application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED) are investigated in this chapter. The aim is to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al [Xu, et al., 1999] is also used here as a comparison with other two proposed methods. These approaches are compared for synthetic lesions at different image signal to noise ratios (SNRs). The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be

quantitatively and more accurately compared. Verification experiments are also performed on clinical skin lesion images.

Chapter 6: This chapter will conclude the final thesis results. Each chapter will be concluded individually and then overall conclusions and suggestions for further work will be made.

References

Binder M, Kittler H, et al., "Computer aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process," *Melanoma Research*, vol. 10, no. 6, pp. 556-561, 2000.

Brenner JF, Lester JM, et al., "Scene segmentation in automated histopathology: techniques evolved from cytology automation," *Pattern Recognition*, vol. 13, no. 1, pp. 65-77, 1981.

Campbell RJ and Flynn PJ, "A Survey of free-form object representation and recognition techniques," *Computer Vision and Image Understanding*, vol. 81, no. 2, pp. 166-210, 2001.

Carli P, De Giorgi V, et al., "Dermatoscopy in the diagnosis of pigmented skin lesions: a new semiology for the dermatologist," *JEADV*, vol. 14, pp. 353-369, 2000.

Cohn-Cedermark G, Mansson-Brahme E, et al., "Trends in mortality from malignant melanoma in Sweden 1970-1996," *Cancer*, vol. 89, pp. 348-55., 2000.

Fisher S and Schmid P, "Analysis of skin lesions with pigmented networks," in ICIP'96, Lausanne, Switzerland., 1996.

Fu KS and Mui JK, "A survey on image segmentation," *Pattern Recognition*, vol. 13, no. 1, pp. 3-16, 1981.

Giachetti A, Frexia F, et al., "Distributed measurement and reporting system for surgical planning," *International Congress Series*, vol. 1256, pp. 828-833, 2003.

Gokmen M, "A comparison on edge detection algorithms based on Gaussian filtering and iteratively refined regularization," in Communication, Control and Signal Processing, Elsevier Science Publishers, 1990.

Goldman L, "Some investigative studies of pigmented nevi with cutaneous microscopy," *J Invest Dermatol*, vol. 16, pp. 407-426, 1951.

Imeliska C, Downes MS, et al., "Semi-automated color segmentation of anatomical tissue," *Computerized Medical Imaging and Graphics*, vol. 24, no. 3, pp. 173-180, 2000.

Jemal A, Devesa SS, et al., "Recent trends in melanoma incidence among whites in the United States," *J Natl. Cancer Inst.*, vol. 93, pp. 678-83, 2001.

Kanzler MH and Mraz-Gernhard S, "Primary cutaneous malignant melanoma and its precursor lesions: Diagnostic and therapeutic overview," *J Am Acad Dermatol*, vol. 45, no. 2, pp. 260-76, 2001.

Kent RO, Kang S, et al., "Clinical diagnosis of pigmented lesions using digital epiluminescent microscopy," *Archives of Dermatology*, 1993.

MacKie RM, Bray CA, et al., "Incidence of and survival from malignant melanoma in Scotland: an epidemiological study," *The Lancet*, vol. 360, no. 9333, pp. 587-91, 2002.

Maintz JBA and Viergever MA, "A survey of medical image registration," *Medical Image Analysis*, vol. 2, no. 1, pp. 1-36, 1998.

Marquez G, "Optical biopsy for the early detection of skin cancer," PhD. Thesis, Texas A&M University, 2001.

Motta G, "Optimization methods for data compression," Ph.D Thesis, Computer Science, Brandeis University, Waltham, Massachusetts, May 2002.

Nixon M and Aguado A, *Feature extraction & image processing*: Butterworth-Heinemann, 2002.

Olabarriaga SD and Smeulders AWM, "Interaction in the segmentation of medical images: A survey," *Medical Image Analysis*, vol. 5, no. 2, pp. 127-142, 2001.

Pehamberger H, Binder M, et al., "In vivo epiluminescence microscopy: improvement of early diagnosis of melanoma," *J Invest Dermatol*, vol. 100:, pp. 356s-362s, 1993.

Shyu M-S and Leou J-J, "A genetic algorithm approach to color image enhancement," *Pattern Recognition*, vol. 31, no. 7, pp. 871-880, 1998.

Steiner A, Pehamberger H, et al., "In vivo epiluminescence of pigmented skin lesions. II. Diagnosis of small pigmented skin lesions and early detection of malignant melanoma," *J Am Acad Dermatol*, vol. 17, pp. 584-591, 1987.

Van der Hoek DP. (2000). Cancer in Australia 1997.[Online]. http://www.sunsmart.com.au/s/facts/skin_cancer_and_australians.htm.

Weinstock MA, "Early detection of melanoma," *JAMA*, vol. 284, no. 7, pp. 886-89, 2000.

Xu L, Jackowski M, et al., "Segmentation of skin cancer images," *Image Visions Computing*, vol. 17, pp. 65-74, 1999. http://www.cs.wright.edu/people/faculty/agoshtas/paper_fig.html

CHAPTER 2: BACKGROUND

Neural networks (NNs) are trying to mimic or simulate the way a simple biological nervous system is believed to operate. The models used by NNs are capable of solving sophisticated, perhaps "intelligent", computation similar to those that the human brain routinely performs. In this chapter NNs are investigated as one of the key elements used in this research. Section 2.2 describes a basic introduction to this field. NNs "learn" from examples, as children learn to distinguish between various objects based on examples that are introduced. There are numerous kinds of NNs and also a wide range of their applications. In this chapter we will, in particular, discuss the applications of NNs to edge detection (section 2.2). The most widely used NNs architecture is chosen in this work and its learning algorithm, Back Propagation (BP) is summarized in section 2.3. The field of automatic image segmentation [Sonka, et al., 1999] is also associated with the work reported in this thesis. Techniques in this field can be classified into four approaches [Fan and Yau, 2001], namely, *1) thresholding techniques, 2) boundary-based methods, 3) region-based methods,* and *4) hybrid techniques*. There is vast number of image segmentation techniques in the literature, however the technique directly related to the work presented in this thesis is addressed (see section 2.4). The mathematical background of optimal thresholding technique is described in section 2.4. Optimal thresholding is the basic building block of the developed iterative segmentation algorithm (IS) and will be discussed in the next chapter.

2.1 Introduction

NNs are designed to simulate the way a simple biological nervous system is believed to operate. Biological neural networks are the local assemblages of neurons (See Figure 2.1) and their dendritic connections that form the (human) brain. They are based on simulated nerve cells or neurons, which are joined together in a variety of ways to form networks (See Figure 2.2). These networks have the capacity to learn, memorize and create relationships amongst data. There are many different types of NNs but some are more popular than others. The most widely used NN is known as the feedforward Back Propagation NNs (Section 2.4). This type of NN uses supervised learning (section 2.2) when transforming input vectors into output vectors and is excellent at e.g. pattern recognition [Jain, et al., 2000]. Another type of NN is the Kohonen or self-organizing network [Kohoen, 1987] that is excellent at finding relationships amongst complex sets of data [Kangas, 1994].

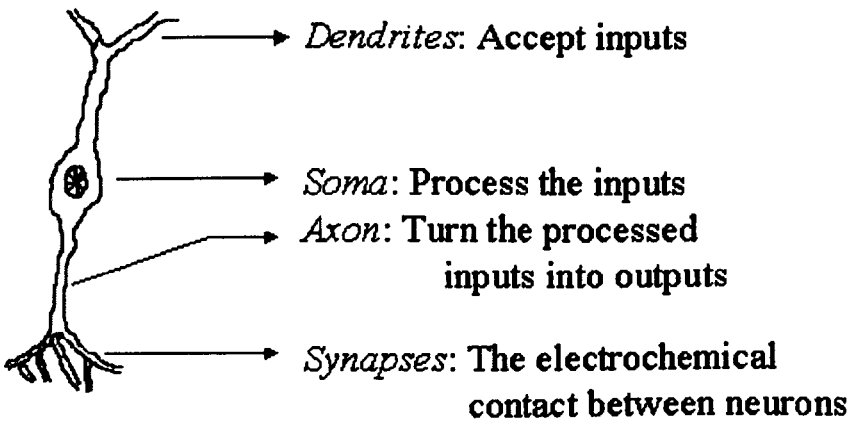


Figure 2.1: Four parts of a typical nerve cell.

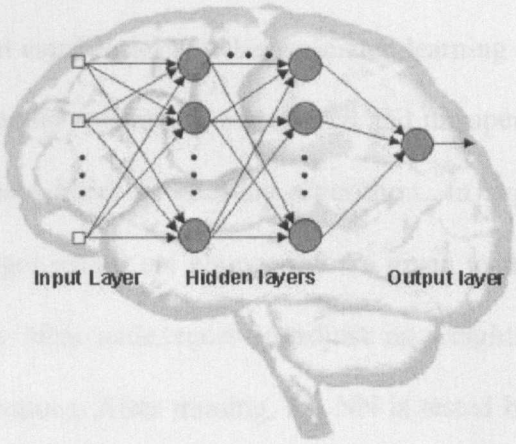


Figure 2.2: Example of feed-forward NNs architecture, which is trying to model billions of biological neurons in our brain.

2.1.1 Learning NNs

Analogous to neurons in the brain the network in a sense learns from examples or by experience just as people do. Neural networks use a set of nodes (processing elements) which are sometimes described as connectionist systems, because of the connections between individual processing nodes. Sometimes the NNs is called as adaptive system or machine learning algorithm, because changing of its connection (training NNs) weights would make the system acquire new knowledge [Haykin, 1994]; each connection between the neurons is stored as a weight-value or strength for the specific connection. Therefore, for the network to learn the solution to a given problem, the values of these connections can change so that the neural network performs more effectively. NNs are also called parallel-distributed processing systems, which give emphasis to the way in which the many nodes or neurons in a neural network operate in parallel [DARPA, 1988].

The Neural networks models differ greatly [Sarle, 2002] . The connection topologies, computational element capabilities and the associated learning algorithms differ from one neural network model to another. Supervised and unsupervised learning [Sarle, 2002] are the two main kinds of learning algorithms. In supervised learning, the desired outputs or target values are known and are given to the NN during training and consequently the NNs endeavours to adjust its weights to try matching its outputs to the target values. After training, the NN is tested by giving it only input values, not target values, and seeing how close it is to outputting the correct target values. In unsupervised learning (e.g. Kohonen or self-organizing network uses unsupervised learning), the NN requires only the input vectors (target values are not provided), which are classified into distinct regions based on regularities and correlations.

2.2 Neural Networks in Edge Detection

A machine vision system often deals with the processing of digital images which consist of a matrix of pixels representing intensities of various positions [Ramalho, 1995]. The main goal of this processing is to provide image improvement for human understanding [Gokmen, 1990;Ramalho, 1995]. This task is achieved by a number of steps. The initial step is the segmentation of the image into meaningful objects [Ramalho, 1995]. Edge extraction is usually the starting step in segmentation because it effectively detects the limits of the objects. These limits are commonly called "edges". Moreover, edge extraction is used mostly in image recognition,

classification or interpretation procedures because it provides a compressed amount of information for processing [Garcya-Silvente, et al., 1997]. The success of an image recognition procedure is related to the quality of the edges marked [Ramalho, 1996].

Neural networks have been successfully applied to pattern recognition tasks and edge detection. This is because of their ability to handle incomplete or corrupted sets of data [Ramalho and Curtis, 1994]. Research in neural networks edge detection have revealed the effectiveness of this technique when applied to real and/or synthetic images. In this section, we are mainly reviewing the existing neural networks edge detection research. However, there is a great necessity to analyse neural network models so as to achieve close insight into their internal functionality. To this purpose, new and general training sets, consisting of a limited number of prototype edge patterns, are proposed to analyse the problem of neural network edge detection (section 3.2, Chapter 3). The high degree of generality of the proposed neural network edge detection method will probably make it more flexible to investigate for further neural network edge detection solutions, which will also meet the needs of certain applications. An example application is presented in Chapters 4 and 5; neural network edge detector NNED has been applied to find sharp edge patterns in noisy synthetic lesion images. *It is interesting to mention here that the previous work in neural network edge detection (listed in Table 1) have limited their analysis and application to certain classes of images whereas in our method the generated set of vectors simply corresponds to standard situations that are unequivocally understood*

as edges or non-edges and permit a controlled distribution of the edges enclosed (section 3.2, Chapter 3).

Table 1 under shows a summary of neural networks edge detectors in different research studies (source is a designation for the author(s) of study). The images used in these studies were either synthetic images (primarily constructed from different edge strengths and orientations [Ramalho and Curtis, 1994]) or real images.

Table 1. Recently published neural network edge detectors.

Author(s)	**Performance presented on**	**Algorithms compared**
[Srinivasan, 1994]	1 synthetic, 1 real	Canny
[Ramalho, 1995]	5 synthetic, 2 real (according to PhD thesis work)	Roberts, Sobel, Prewitt, Frei, Canny, Deriche, O'Gormann.
[Leow and Lua, 1997]	2 real	Boundary Contour System (BCS) and Boundary Segmentation Network (BSN)
[Wong, et al., 2000]	4 real	Shen-Castan
[Rajab and Woolfson, 2003]	Standard edge patterns (prototype patterns)	Sobel

Ramalho [Ramalho, 1996] compared the capability of a neural network for edge detection with different conventional edge detection methods. Moreover, the study also established a strategy to arbitrate between any two edges maps (for some selected cases such as Roberts versus Canny filters) using the neural network. The arbitration system is taught from data sets that are extracted from three edge maps.

These maps include the two arbitrated edge maps and the reference edge map that could be generated from a synthetic image [Ramalho, 1995]. Inspired by the parallel nature of the neural network arbitration system, the number of algorithms that are used in the arbitration process was kept to two; note that an increase in the number of methods to be arbitrated would also increase the duration of network learning phase. Srinivasan [Srinivasan, 1994], Leow [Leow and Lua, 1997], and Wong [Wong, et al., 2000] each proposed a neural network model for edge detection and, in contrast to Ramalho [Ramalho, 1995], neural network arbitration was not considered.

However, to recognize objects in an image we also need to perform *boundary segmentation* [Leow and Lua, 1997]. Several neural networks methods proposed in the literature, not only perform edge detection but also boundary segmentation, i.e., to identify the edges that outline the object's boundaries. An improved version of the Boundary Contour System (BCS) [Grossberg, 1985], called the Boundary Segmentation Network (BSN) is able to extract thinner edges (in contrast with BCS) and is better at rejecting noisy background details, when segmenting the object's overall boundaries. Both of the two networks consist of five layers of units. In the BSN, the last sharpening layer performs the task of thinning edges, which is absent in the original work of BCS. Large networks like these can be particularly worthwhile if speed and cost are not important factors. Careful design of large architecture and specification of the tasks required by each layer are considered in these topologies. This could avoid the major drawbacks of large networks that they significantly learn more patterns; not only do they learn the main characteristics of

the learning set but also the particularities of that set [Ramalho, 1996]. As a consequence of these large networks, they become an associative memory with little generalization capabilities.

One of the aims of this thesis is to investigate the application of neural network edge detection (NNED) to real and synthetic noisy images. *However, the NNED classification between edges and noise patterns is a critical issue especially for the case of very noisy images. In this case, the inputs to the NNED are the original noisy patterns, and if we also considering that there are no preprocessing steps that are used such as noise filtering operations that occur in non-neural network algorithms, then using the NNED will be advantageous to reduce the amount of image processing required and to minimize the execution time.* In chapter 4, the NNED method is considered in more detail.

2.3 MLP and Back Propagation Algorithm

The Multi Layer Perceptron (MLP) is one of the most common neural network architectures that has been used successfully in various applications. The Back propagation algorithm is normally used to train these networks. The MLP is a development of a Single Layer Perceptron (SLP) from the simple Delta rule. The rule is based on the idea of continuously modifying the strength of the input connections to reduce the difference (the delta) between the desired target value (output) and the current actual output of a neuron.

The error back-propagation algorithm [Rumelhart, et al., 1986] is used for training MLP feedforward neural networks to minimize the training error [Nakashima, et al., 2001]. The network topology is constrained to be *feedforward*: i.e. loop-free. After a neuron performs its function it passes its output to all of the neurons in the layer below it, providing a feedforward path to the output. As shown in the Figure 2.2, MLPs consists of one input layer, one output layer, and the possibility of one or more hidden layer(s).

Initially the network's weights are initialized to small random values, and the feed-forward operation is performed on a presented input pattern.

$$y_j^l = f(\sum_{i=1}^{N_{l-1}} w_{ij}^l x_i^l) \tag{2.1}$$

where N_{l-1} is the number of neurons in the previous layer and l represents the layer number. The sigmoid function (Eqn 2.2), as an activation function, is often used to produce a nonlinear approximation function. The advantage here is to have a simple derivative that is needed during the weight adaptation.

$$y = f(x) = \frac{1}{1+e^{-x}} \tag{2.2}$$

$$y' = y(1-y) \tag{2.3}$$

Rumelhart et al. [Rumelhart, et al., 1986], define an error measure for the output layer weights, which is the difference between the calculated output (y) and the desired output vector (d) or the target output value.

$$\delta_j^L = f'(y_j^L)(d_j - y_j) \tag{2.4}$$

Since the hidden layers have no explicit target values, their error measure is obtained by back-propagating the output errors,

$$\delta_j^l = f'(y_j^l)\sum_k \delta_k^{l+1} w_{jk}^{l+1} \tag{2.5}$$

The synaptic weight for all layers are updated using these error values,

$$w_{ij}^l(t+1) = w_{ij}^l(t) + \Delta w_{ij}^l(t) \tag{2.6}$$

$$\Delta w_{ij}^l(t) = \eta \delta_j^l(t) x_i^l + \alpha(w_{ij}^l(t) - w_{ij}^l(t-1)) \tag{2.7}$$

where t represents the current iteration, η is the learning rate and α is the momentum rate.

2.4 Optimal thresholding

The main aim in image segmentation is to divide the image into specific regions from which we can extract appropriate features for analysis [Hintz-Madsen, et al., 1999]. Several techniques have been reported for image segmentation and the most commonly used has been colour or greyscale thresholding [Gao, et al., 1998]. The assumption made in the thresholding technique is that adjacent pixels (whose value lies within a certain range e.g. grey level, colour value, texture, etc) belong to the same class [Fan and Yau, 2001;Lim and Lee, 1990]. Good segmentation of images can be obtained by thresholding techniques, if a segmented image includes only two opposite components [Fan and Yau, 2001]. The reader should refer to [Sonka, et al., 1999] for a review of thresholding techniques.

In this section we will present the theoretical background of an optimal threshold [Hintz-Madsen, et al., 1999] to segment an image according to its luminance levels of pixels. It should be noted that the iterative scheme [Ridler and Calvard, 1978] was used by [Hintz-Madsen, et al., 1999] to estimate the $N-1$ optimal thresholds and the N luminance means.

Let us denote the image as, R, and N regions of this image as, R_i, $i=1,2,...,N$. This may be formalised as

$$R = \bigcup_{i=1}^{N} R_i, \quad R_i \cap R_j = \phi \ , \ i \neq j \tag{2.8}$$

Here, regions R_i differ from each other according to some chosen property like, e.g., luminance, colour, ...etc. The derived optimal threshold T_i is desired to separate pixels in region R_i from pixels in regions R_{i+1}. The luminance levels are considered and analysed in our case in order to determine what region a pixel belongs to. If we consider the luminance level, $l(m,n)$, to be a sample of a stochastic variable, l, and assuming we know the conditional probability distribution $p(l|R_i)$ and the prior region probability $p(R_i)$, then the problem can be viewed (of selecting these thresholds) as a classification problem. Therefore, the Bayesian rule to minimize the probability of a misclassification (e.g. between regions R_i and R_{i+1}) is

$$p(R_i)\, p(l \mid R_i) = p(R_{i+1})\, p(l \mid R_{i+1}) \tag{2.9}$$

Let us assume the two density functions of $p(l|R_i)$ and $p(l|R_{i+1})$ are Gaussian with mean μ_{R_i} and $\mu_{R_{i+1}}$ respectively and equal variance σ^2. The two density functions will be

$$p(l \mid R_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\left(l - \mu_{R_i}\right)^2 / 2\sigma^2\right] \tag{2.10}$$

and

$$p(l \mid R_{i+1}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\left(l - \mu_{R_{i+1}}\right)^2 / 2\sigma^2\right] \tag{2.11}$$

Substituting Eqns (2.10) and (2.11) in Eqn (2.9), we obtain the following result:

$$\frac{\exp\left[-(l - \mu_{R_i})^2 / 2\sigma^2\right]}{\exp\left[-(l - \mu_{R_{i+1}})^2 / 2\sigma^2\right]} = \frac{p(R_{i+1})}{p(R_i)} \tag{2.12}$$

Taking logarithms of Eqn (2.12)

$$\left(l - \mu_{R_i+1}\right)^2 - \left(l - \mu_{R_i}\right)^2 = 2\sigma^2 \log \frac{p(R_{i+1})}{p(R_i)} \tag{2.13}$$

The desired region R_{i+1} is thus defined by all luminance values $l(m,n)$ corresponding to

$$l(m,n) > \frac{\mu_{R_i} + \mu_{R_{i+1}}}{2} + \frac{\sigma^2}{\mu_{R_i} - \mu_{R_{i+1}}} \log \frac{p(R_{i+1})}{p(R_i)} \tag{2.14}$$

or the optimal threshold T_i between regions R_i and R_{i+1} is

$$T_i = \frac{\mu_{R_i} + \mu_{R_{i+1}}}{2} + \frac{\sigma^2}{\mu_{R_i} - \mu_{R_{i+1}}} \log \frac{p(R_{i+1})}{p(R_i)} \tag{2.15}$$

Assuming the prior probabilities $p(R_i)$ and $p(R_{i+1})$ are equal, then equation (2.15) reduces to

$$T_i = \frac{\mu_{R_i} + \mu_{R_{i+1}}}{2} \tag{2.16}$$

It is interesting to mention here that an early step in the analysis of dermatoscopic imagery is the separation of lesion from background [Hintz-Madsen, et al., 1999]. The optimal thresholding algorithm has been used by [Hintz-Madsen, et al., 1999] and has provided good results when it was applied to segment all the tested dermatoscopic images. In the algorithm presented in section 3.3.2, chapter 3, we will use an optimal threshold value to segment the entire region containing the lesion from its surrounding skin.

2.5 Conclusion

In this chapter the theoretical background and a literature review are presented of the application of Neural Networks (NN) to edge detection. NNs are trying to mimic or simulate the way a simple biological nervous system is believed to operate. The models used by NNs are capable of solving sophisticated, perhaps "intelligent", computation similar to those that the human brain routinely performs. NNs are

investigated as one of the key elements used in this research. There are numerous kinds of NNs and also a wide range of their applications. Discussion of the applications of NNs to edge detection is particularly presented. The most widely used NNs architecture was chosen in this work and its learning algorithm, back propagation (BP) training algorithm. Finally, the mathematical background of another edge detection technique, optimal thresholding, is described which is the basic building block of the developed IS algorithm.

References

DARPA, *Neural network study*: AFCEA International Press, p. 60, 1988.

Fan J and Yau DKY, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1454-1466, 2001.

Gao J, Zhang J, et al., "Segmentation of dermatoscopic images by stabilized inverse diffusion equations," in Proceedings of the International Conference on Image Processing '98, IEEE, Piscataway, NJ, 1998, vol. 3, pp. 823-827.

Garcya-Silvente M, Garcya JA, et al., "A new edge detector integrating scale-spectrum information," *Image and Vision Computing*, vol. 15, pp. 913-923, 1997.

Gokmen M, "A comparison on edge detection algorithms based on Gaussian filtering and iteratively refined regularization," in Communication, Control and Signal Processing, Elsevier Science Publishers, 1990.

Haykin S, *Neural networks: A comprehensive foundation*: NY: Macmillan, 1994.

Hintz-Madsen M, Hansen LK, et al., "probabilistic neural network framework for detection of malignant melanoma," Technical University of Denmark, Dept. of Mathematical Modeling, Lyngby, Denmark, Technical report 1999.

Jain AK, Duin RPW, et al., "Statistical pattern recognition: A review," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, 2000.

Kangas J, "On the analysis of pattern sequences by self-organizing maps," PhD. Thesis, Helsinki University of Technology, Finland, 1994. http://citeseer.nj.nec.com/cs

Kohoen T, "Adaptive, associative, and self-organising function in neural computing. Artificial neural networks," *Applied Optics*, vol. 26, no. 23, pp. 4910-4918, 1987.

Leow WK and Lua SC, "An improved neural network for segmenting objects' boundaries in real images," in Proceeding of International Conference on Neural Network, 1997, vol. III, pp. 1663-1668.

Lim YW and Lee SU, "On the color image segmentation algorithm based on the thresholding and the fuzzy C-means technique," *Pattern Recognition*, vol. 23, no. 9, pp. 935-952, 1990.

Nakashima A, Hirabayashi A, et al., "Error correcting memorization learning for noisy training examples," *Neural Networks*, vol. 14, pp. 79-92, 2001.

Rajab MI and Woolfson MS, "Analysis of neural network edge detection," in WSEAS Transactions on Systems; 4th WSEAS International Conference on Neural Networks and Applications (WSEAS NNA 2003), WSEAS, Vouliagumeni,Athens, Greece, 2003, vol. 2 Issue(3), pp. 649-654.

Ramalho M, "Edge detection using neural network arbitration," PhD thesis, School of Elect. and Electron. Eng., University of Nottingham, Nott., UK, 1996.

Ramalho M, "Edge Detection Using Neural networks: A Comparison," in Int. Conf. on DSP'95, Limassol, Cyprus, 1995, pp. 596-601.

Ramalho M and Curtis KM, "Neural Network Arbitration of Edge Maps," in Transputer Applications and Systems '94, IOS Press, Netherlands, 1994, pp. 32-39.

Ridler TW and Calvard S, "Picture thresholding using an iterative selection method," *IEEE Trans. on Systems, Man and Cybernetics*, 1978.

Rumelhart DE, Hinton GE, et al., "Learning internal representations by error propagation," in Parallel Distributed Processing: Explorations in the Microstructures of Cognition, MIT press, Cambridge, MA, 1986, pp. 318-362.

Sarle WS. (2002). How many kinds of NNs exist? [Online]. Available e-mail: saswss@unx.sas.com. FAQ Neural Networks: ftp://ftp.sas.com/pub/neural/FAQ.html.

Sonka M, Halavac V, et al., *Image Processing Analysis and Machine Vision.* London, UK: Chapman & Hall, 1999.

Srinivasan V, "Edge Detection Using Neural Network," *Pattern Recognition*, vol. 27, no. 12, pp. 1653-1662, 1994.

Wong H, Caelli T, et al., "A model-based neural network for edge characterization," *Pattern Recognition*, vol. 33, pp. 427-444, 2000.

CHAPTER 3: METHODOLOGY

In this chapter the three techniques that are applied in this research to the segmentation of pigmented skin lesions are described. We investigate the application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED). The aim is to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al. [Xu, et al., 1999] is also used here to compare with the other two proposed methods. These three approaches will be compared for synthetic lesions at different image signal to noise ratios (SNRs) as discussed in Chapter 5.

3.1 Data description

3.1.1 Observations

Skin lesions exhibit a variety of colours. The presence of this variety of colours is not very useful in segmenting images [Xu, et al., 1999]. However, one can observe that there are changes in colour when going from a lesion to its background (its surrounding healthy skin). These changes are similar in most skin images and can be used to simplify, effectively, the segmentation of skin lesion images. Furthermore when we notice colour variations inside and outside the lesion then a second observation can be carried out. Since the delineation of a skin lesion deals with the boundaries of that lesion then colour changes belonging to the lesion boundaries are

all necessarily important in this process. Therefore, we can ignore both colour changes inside a lesion and in the background of the lesion. [Xu, et al., 1999] stated two main facts which are relevant to most skin lesion images and can be summarized as follows:

1. Skin lesions come in a variety of colours, however, changes in color from a lesion to its background (its surrounding skin) are similarly observed in all images.
2. Skin lesions exhibit significant color variations, which may exist within a lesion or in the background.

These cases can be simulated by producing synthetic binary images of the lesions. After adding Gaussian noise, the resulting lesion intensities will vary between lesion and background and also exhibit variations within this lesion or in the background. Clearly the monochrome noisy images neglect color that is present in optical images.

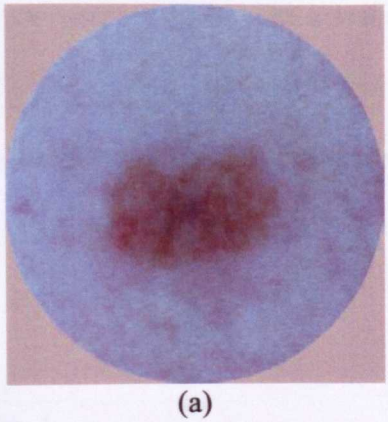
Gaussian noise is chosen as a first approximation noise source with which to compare the different techniques (Chapter 5). Real lesion images may contain noise such as details from skin texture and hair that make it more difficult to localize the lesion border. In addition, the effect of color variations such as shadows and reflections (bright spots) tend to bias the color map when performing color segmentation. However, modeling these effects is beyond the scope of this study.

3.1.2 SIAscope images

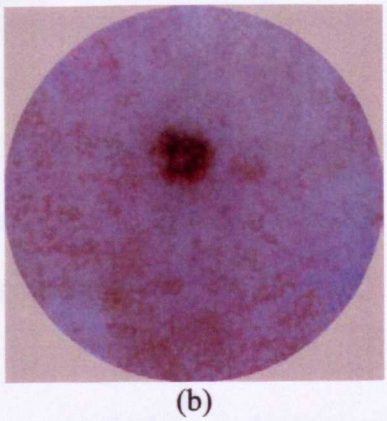
The medical data used in this research are mainly based on the images supported by Astron Clinica Limited, UK. Figure 3.1 shows three high resolution colour skin lesions which are captured by SIAscope's camera. They are SIAscope images; SIA stands for "Spectrophotometric Intracutaneous Analysis" which is a rapid dermal scanning technique that uses remitted light in the visible and infrared spectra to gain information concerning the skin [Satta and Claridge, 2002]. The SIAscopy is the result of many years' joint development by Astron Clinica Limited, The University of Birmingham Computer Science Department and Addenbrooke's Hospital, Cambridge, and has been shown to have significant advantages in certain areas of dermatology. The detailed theory of SIA has been presented in previous papers [Cotton, et al., 1997;Cotton, et al., 1999]. The clinician is presented with SIAgraphs that are high resolution images demonstrating information on the distribution, location and quantity of various chromophores and other structures throughout a lesion. Furthermore, in the diagnosis of malignant melanoma, a clinical study using the SIAscope has identified a number of features in SIAgraphs which are correlated with underlying significant histopathology [Moncrieff, et al., 2002;Satta and Claridge, 2002].

3.1.3 Synthetic lesions

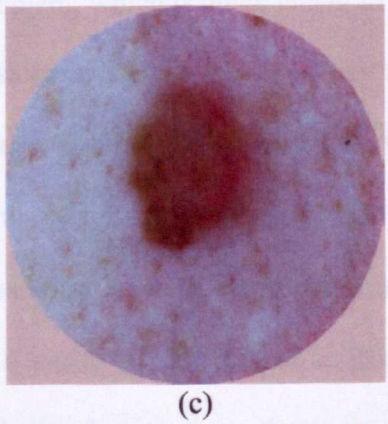
It is worth mentioning here the importance of irregularity of the lesion border, as it is one of the important clinical features that aid in differentiating benign melanocytic nevi from malignant melanoma. There has been recent work in the development of



(a)



(b)



(c)

Figure 3.1: High resolution SIAscope colour images showing three skin lesions.

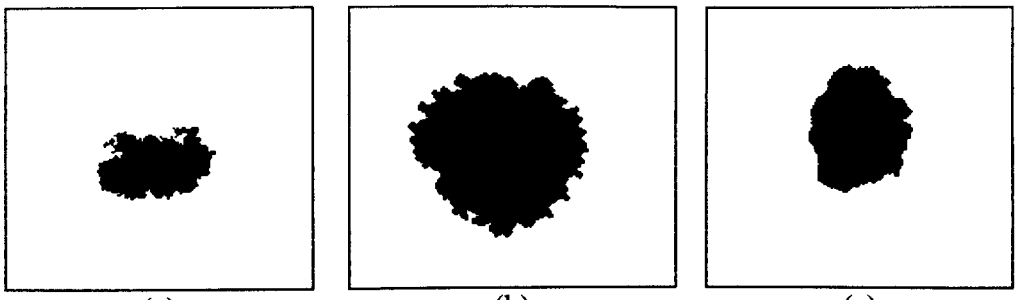
more sophisticated techniques to measure border irregularity, such as the method developed by [Lee, et al., 2003], which could offer extra features such as localization of the significant indentations and protrusions in a lesion. A measure like this is sensitive to structure irregularity and may be more useful for accurate classification of the lesion shape. Amongst numerous irregularity measures, the border irregularity used here [Gray, 1971;Rosenfeld, 1977] is given by:

$$I = \frac{P^2}{4\pi A} \tag{3.1}$$

where P is the perimeter of the segmented lesion and A is its area. For example, and according to this measure, a circular border has a unity irregularity ($I = 1$), which reveals a perfectly regular border.

To build an analysis environment, which is much closer to the real situation, we have used the three real skin lesions shown in Figure 3.1 as the base to generate three corresponding typical synthetic lesions (See Figure 3.2). Golston et al. [Golston, et al., 1992] , determined by the use of some preliminary test images, that a threshold of $I = 1.8$ makes the best separation between regular and irregular borders, that is, a lesion with I greater than 1.8 was to be classified as an irregular lesion. According to this classification threshold, the three lesions shown in figure 3.2 represent three types of lesion topology according to their border irregularity indices (Eqn. 3.1), these are a) irregular lesion I=2.8, b) regular lesion I=1.8 and c) very regular lesion

I=1.1. We can say that these three lesions are chosen as a basic typical test lesions, which would represent three candidates of border irregularities at two extremes of the threshold of 1.8 and at the threshold itself.



(a) (b) (c)

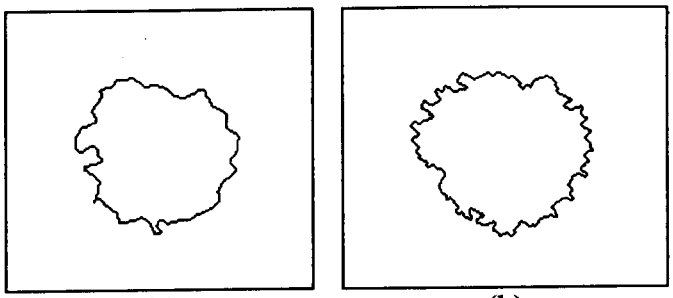
Figure 3.2: Three cases of synthetic lesions: (a) Irregular I=2.8, (b) regular I=1.8, (c) Very regular I=1.1

It should be noted that accurate segmentation methods such as the IS method, which is developed in this work, could provide accurate lesion boundaries (see Chapter 5). For this reason we have adopted this method to support generating the synthetic lesions shown in Figure 3.2. Moreover, an accurate method like the IS could accurately delineate a boundary of a lesion which represents the structure irregularity of that lesion and may also suggest excessive cell growth or regression of a melanoma.

3.1.3.1 Formation

The three binary synthetic lesions shown in figure 3.2 have been formed using the IS method presented in Section 3.3. This is because this method is much more flexible than the other two methods (NNED and ASS methods) to generate binary segments

with a wide range of structure irregularities. However, there might be a doubt that any expected results using these synthetic lesions would be biased towards the IS method. To avoid this, we have added manually an additional complexity to the border of one of the three lesions (figure 3.2(b)) so that the final border structure differs from that determined by the IS method, as demonstrated in figure 3.3, and can also be considered as if it is generated by another method.



(a) (b)

Figure 3.3: (a) IS segmentation result of lesion shown in figure 3.1(b) with irregularity index $I = 1.3$. (b) Border in (a) is manually modified such that $I = 1.8$.

In chapter 5 we will compare the robustness of segmentation errors to noise for the three segmentation methods (IS, ASS, and NNED) using these lesions and after being corrupted by different realizations of Gaussian noise. Since we know that segmentation methods may yield slightly different errors when applied to different lesion structures, then it is tedious and time consuming to generate numerous similar structures of synthetic binary lesions as those shown in figure 3.2. We have instead

taken three references synthetic images representing typical lesions with very different irregularity indices. Adding an effect to these binary lesions such as the Gaussian noise investigated in our study, will give an indication as to the robustness of the performance of each segmentation method to noise – this will be discussed later in Chapter 5.

3.2 Automatic skin segmentation (ASS) method

The ASS method is chosen as a benchmark with which to compare the IS and NNED methods as its potential has already been established in previous studies and it is available in the public domain. This method has been described in detail elsewhere [Xu, et al., 1999] and so will only be summarized briefly here. The segmentation method was developed in three steps: preprocessing, initial segmentation, and region refinement. The preprocessing step of the ASS segmentation method starts with the transformation of a colour image into an intensity image, and mapping it using a Gaussian function. The mapping function reduces image gradients in the input that correspond to details in the background; the enhancement process is considered as the main contribution of this work. Consequently, initial segmentation is applied to the obtained enhanced intensity image to find an approximate optimal lesion boundary. An initial threshold value is first determined based on the observation that highest gradients of pixels are generally more probable on lesion boundaries than inside or outside lesions. Therefore, the average intensity of the top $p\%$ highest gradient pixels in the obtained intensity image is used to compute an initial threshold value (T). The initial threshold value is used to find approximate lesion boundaries.

Any error made in this initial segmentation will be corrected in the *region refinement*: a mask is defined between two thresholds (T_1 and T_2) and inside which the approximate lesion boundary is expected to exist. Selected sample outputs of each processing stage [Xu, et al., 1999] are demonstrated in Figure 3.4.

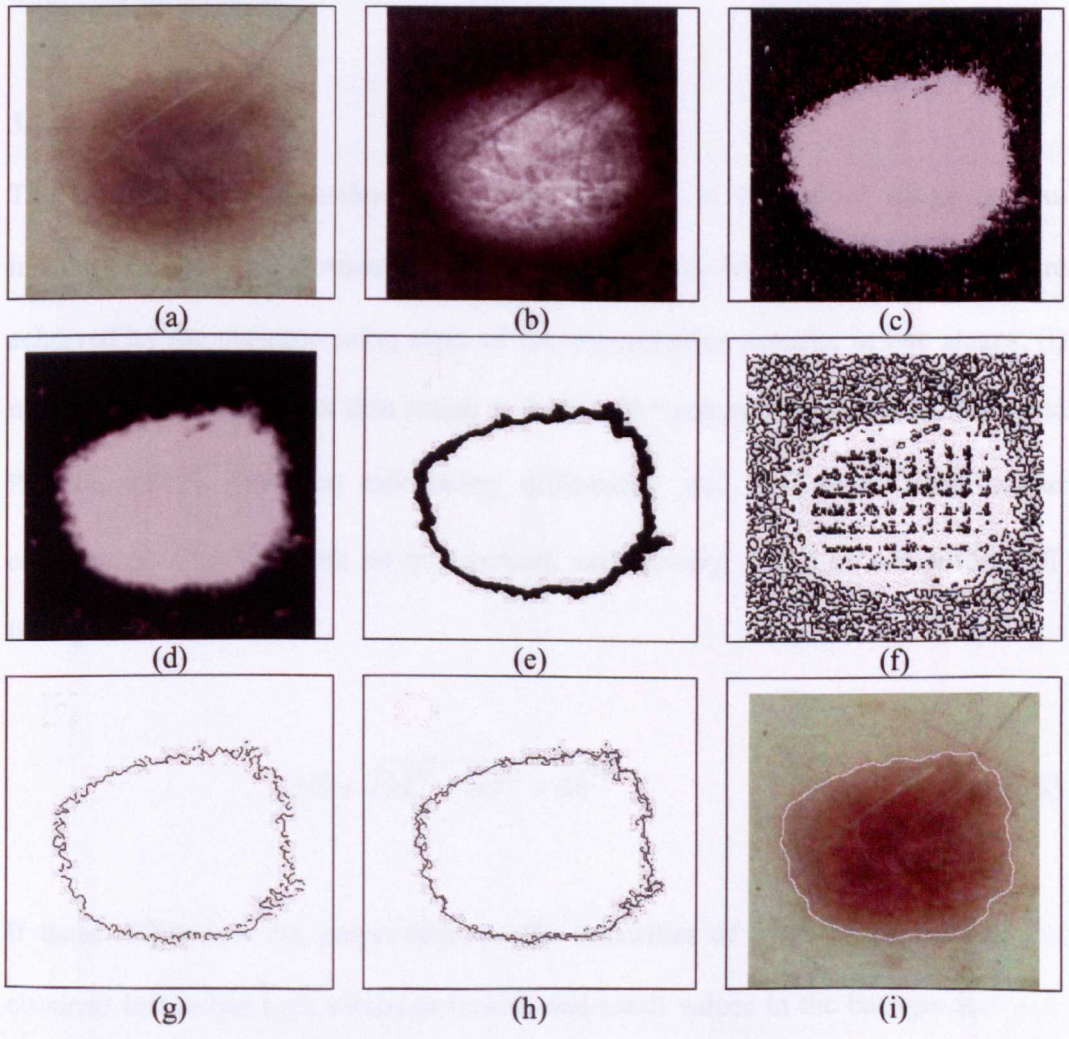


Figure 3.4: (a) Original colour lesion. (b) Intensities obtained after mapping colours of image according to Eqn. 3.2. (c) Mapping intensities according to Eqn. 3.4. (d)

Image in (c) is smoothed with a 2D Gaussian kernel of standard deviation 2 pixels. (e) Image obtained from double thresholding step. (f) Authentic edges of (d). (g) Edges of (f) falling in the mask of (e). (h) Edges obtained after fitting an initial contour. (i) Example of the final segmentation result when using $p = 8\%$, $d = 10$ pixels and $\sigma = 3$ pixels. *Images from [Xu, et al., 1999].*

The three major processing steps of this method are also summarized briefly in the following subsections:

3.2.1 Preprocessing

The first step in this method is the transformation of the colour image into an intensity image. Transforming colour images of skin lesions into intensities are achieved by the pre-processing stage of the segmentation process, in two stages: (i) mapping colour images of skin lesion to the L*a*b* colour coordinates [Kasson and Plouffe, 1992], (ii) then calculating differences $\Delta L^*, \Delta a^*, \Delta b^*$ of each colour components L^*, a^*, b^* with its background, respectively, and then computing ΔE from [Hill, et al., 1997]:

$$\Delta E = \sqrt{\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}} \tag{3.2}$$

If these differences are proportional to the intensities of the background then the obtained image has high values in lesions and small values in the background or in an other word the lesions will be shown as bright spots. This is demonstrated in Figure 3.4(b). After assuming that the lesion does not fall on image corners, Xu. et

al. estimate the lesion background for each colour components L^*,a^*,b^* from their median color instead of using the average colour. This is because in the case of the occurrence of peak color pixels between the skin colour pixels, such as hair color pixels, then the median value would exclude these peak colours and hence only the skin color pixels are used to estimate the color of the background. In this estimation a small window of 10 ×10 pixels in size was taken from the four corners of an image.

Since our interest is in changes from the background to a lesion or vice versa then suppressing significant colour variations inside a lesion or in the background are necessary. To implement this then a function that provides the property shown in Figure 3.5 is needed.

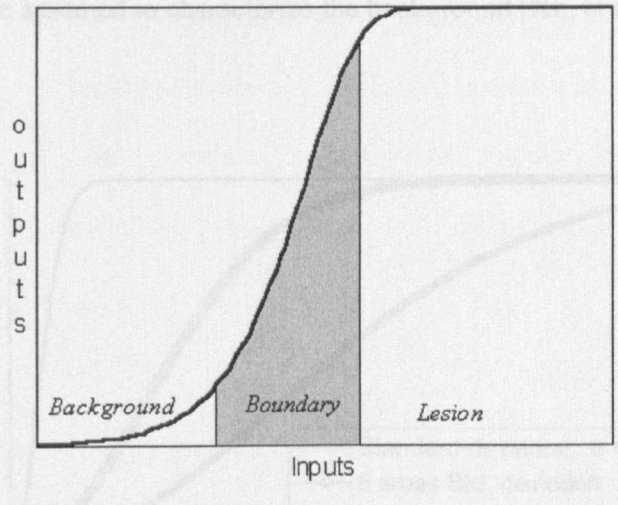


Figure 3.5: Property of a desirable mapping function.

Therefore, the function is trying to reduce image gradients corresponding to details belonging to the background and to a lesion and, at the same time, the gradients of

the intensities of pixels falling on lesion boundaries are increased (the gradients of boundary region are increased as depicted in the shaded region of Figure 3.5). Xu et al. approximated this function using the following Gaussian function *F(x)* computed from:

$$G(x) = 1/\sqrt{2\pi\sigma} \quad e^{-x^2/2\sigma^2} \tag{3.3}$$

and

$$F(x) = 1/\sqrt{2\pi\sigma} \quad -G(x) \tag{3.4}$$

The standard deviation of intensities in the four windows (taken from the four corners of an image) is used as the standard deviation in function *F(x)*, since these four corners are assumed to characterize the background [Xu, et al., 1999].

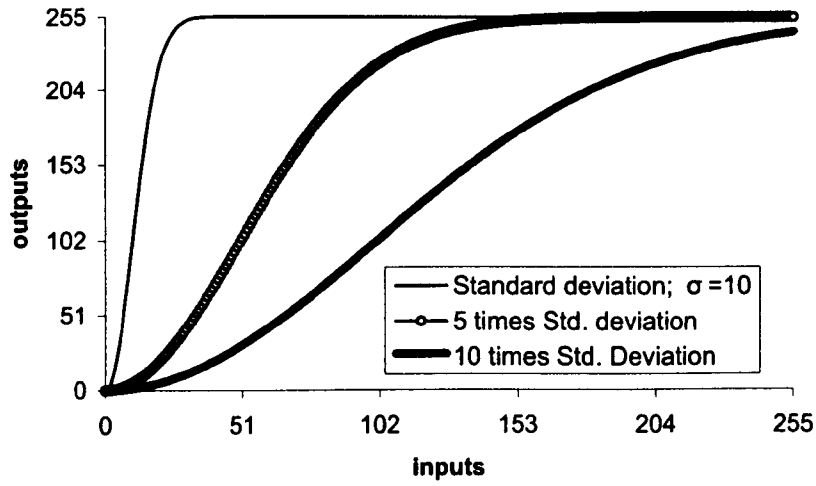


Figure 3.6: Gaussian mapping function $F(x)$ (Eqn. 3.4) at three different

standard deviations: σ =10, 50, and 100.

3.2.2 Double thresholding

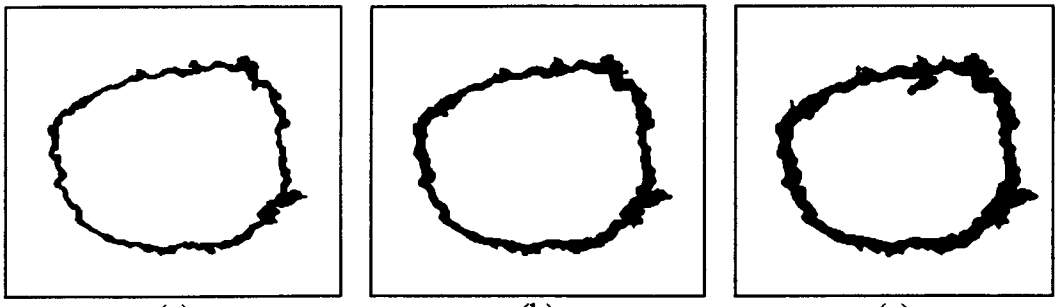
Double thresholding is used here to focus in on an image area where the optimal position and shape of a lesion boundary is likely to exist. The two thresholds are determined in such a way that

$$d \times n = \sum_{i=T_1}^{T-1} h[i] = \sum_{i=T+1}^{T_2} h[i] \tag{3.5}$$

where $h[i]$ is the number of image pixels with intensity, $i = 0,...,255$, and T is automatically determined by thresholding the image at the average intensity of the $p\%$ highest gradient pixels in the image; p is typically a small number, e.g. 5. A second input parameter of the segmentation method along with $p\%$, is the parameter d which is the average displacement that is needed to approximate n points (n is also one of the input parameters proposed by this method) found in the region boundaries at threshold T. Therefore the product $d \times n$ will approximate the total number of boundary points to obtain the optimal boundary.

Two assumptions are made in this method: (i) a point on the lesion boundary is obtained by intensity thresholding T needs to be displaced by d pixels so as to fall on the optimal boundary pixel and (ii) the average displacement can not exceed d pixels; we can say that d represents the average correction anticipated of boundary pixels when thresholding the image at intensity T. Parameter d is dependent on the gradient

at the pixel: a larger d would be necessary if the pixel has low gradient and vice versa. The sensitivity of this parameter is obvious when the gradient is low and any slight change to the threshold value T would move the boundary point more than when the gradient is high. As a result of this fact, when d is too small some parts of the optimal boundary may be missed, while choosing too large a value for d makes the detection of the optimal boundary difficult. Figures 3.7 below shows three examples of double thresholding when applied to a smoothed intensity lesion by setting the parameter d = 5,10, and 15, respectively. It is clearly shown that the growth of the boundary region into neighbouring regions (where optimal boundary pixels are less likely to exist) is proportional to the increase of parameter d.



(a) (b) (c)

Figure 3.7: Double thresholding the image shown in Figure 3.4(c) at: (a) d=5, (b) d =10, and (c) d=15.

3.2.3 Region refinement

In the presence of noise, the boundary region potentially merges with another boundary so the mask needs to be refined. Combining image edge information (Figure 3.4(f) shows the authentic edges [Clark, 1989] that represents the locally

maximum gradient pixels of Figure 3.4(d)) with a double thresholded mask, and then focusing on a small fraction of image edges that lie on or near the initial lesion boundary can achieve this refinement, see figures 3.3(e)-(h). Since the obtained edges may not form a closed contour, the contour segments must be connected to construct a closed contour at the approximate boundary. The approach chosen uses a rational Gaussian curve fitting [Goshtasby, 1995] as a closed elastic curve to shrink or expand to fit the edges in its neighborhood.

3.3 Iterative segmentation (IS) method

In this section the first developed approach to segment a skin lesion will be introduced. The principle aim in the evaluation of a segmentation technique is to study the effectiveness in distinguishing a lesion from its surrounding background. According to that fact, two criteria are considered in this method:

- an accurate search of the optimal lesion segment can be achieved by analyzing the whole image and consequently the true lesion border is also retrieved. Therefore the need to approximate the lesion border within the refined regions (Section 3.2) will be eliminated. In another word the curve fitting approach is omitted in this method.
- Input parameters mostly are image dependent (e.g. d and n in ASS method), and have to be adjusted based on the properties of the class of images used in the segmentation method. It is also desired to minimise number of input parameters so that the user is required to provide as low as possible number of input parameters to segment a skin lesion image.

Three steps are used in this method to segment the noisy synthetic lesions; intensity mapping, application of an isodata algorithm, followed by object outlining. The use of synthetic images in the design and development stage has a great importance to facilitate and ease the design process, since, by knowing the exact locations of lesion boundaries, initial comparisons can be made to assess the potential of the suggested method. Figure 3.8 demonstrates a test experiment showing various outputs of imaging tasks of the iterative method which will be discussed in subsections 3.3.1 to 3.3.3. The synthetic lesion shown in Figure 3.2(b) is selected to show the robustness of this method to retrieve the true border of the original skin lesion.

The segmentation tasks are expected to work on intensities lesion images (similar to image shown in figure 3.4(b)). The three major processing steps of this method are explained in the following subsections:

3.3.1 Intensity mapping

A mapping function $F(\Phi)$ is used here to map the intensities I to enhance features at the boundary;

$$F(\Phi) = k\left(1 - \exp\left(-\Phi^2 / 2\sigma^2\right)\right) \tag{3.6}$$

where

$$\Phi(I) = c^2 I^3 + c^4 I^5 + c^5 I^7 \tag{3.7}$$

and

$$c = 1/k \tag{3.8}$$

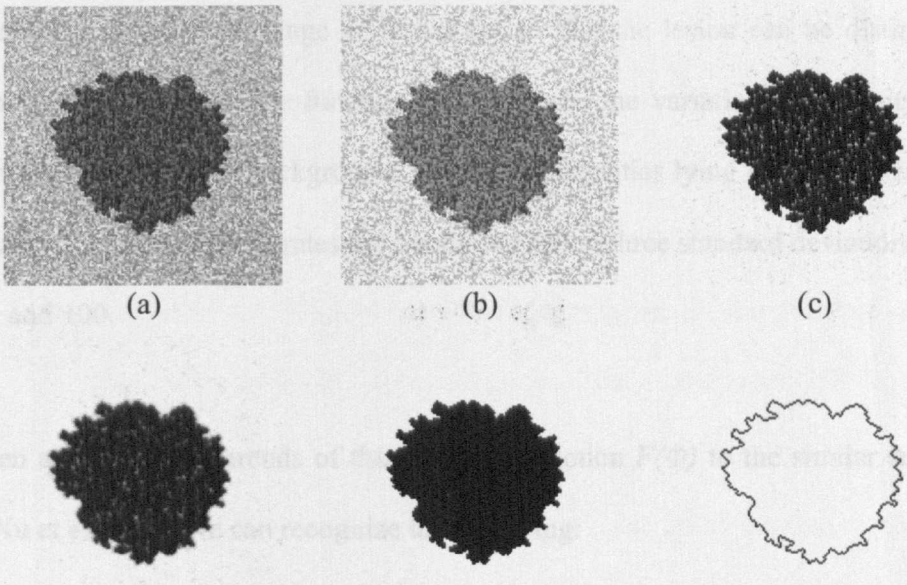


Figure 3.8: Iterative segmentation process: (a) Noisy synthetic lesion (S/N=3.0), (b) subtraction of median background, (c) intensity mapping by function $F(\Phi)$ (Eqn. 3.2), (d) smoothing by a 2D Gaussian kernel of size 3x3 and standard deviation 0.6 pixel, (e) processing by isodata algorithm to produce a binary segment at an optimal threshold, (f) final segmentation result by outlining to produce one pixel wide edges.

c is a scaling constant ($c<1$), and σ is the standard deviation of the Gaussian. We use in our implementations the range of intensities of I between 0 and 255 and hence the normalization factor k will be 255. $F(\Phi)$ achieves less redundancy in the color map than the Gaussian transformation used in [Xu, et al., 1999] which makes it more suitable to map a wide range of intensities so that the lesion can be distinguished from the background. The function $F(\Phi)$ reduces the variation in intensity within both the lesion and the background. However, intensities lying on the border will be enhanced. Figure 3.9 illustrates the trends of $F(\Phi)$ at three standard deviations σ =10, 50, and 100.

When comparing the trends of this mapping function $F(\Phi)$ to the similar suggested by Xu et al. ($F(x)$) we can recognize the following:

(i) $F(x)$ has a major drawback when an image contains low variations in its background, which results in a low value of the standard deviation being used in $F(x)$. Unlike the desired function shown in Figure 3.5, $F(x)$ will map intensities such that both the background and boundary gradients are increased. According to that, the function will include details of the background in the resulting mapped image. We can conclude that this mapping function is only able to work properly according to the property of the desired function shown in Figure 3.5, on a class of images containing significant variation in intensity in their background such as noise that appears in the most noisy clinical colour lesion images. It

should be noted that discussions of results on a wide range of images will be presented in Chapter 5.

(ii) $F(\Phi)$ has a major advantage when an image contains low variations in its background again resulting in a low standard deviation being used in $F(\Phi)$. Compared with $F(x)$ shown in Figure 3.4, $F(\Phi)$ in Figure 3.9 will map intensities such that both gradients of the background and inside of the lesion are decreased. Therefore, the function will suppress details of the background in the resulting mapped image. A similar behaviour of the function $F(\Phi)$ is also revealed at high values of the standard deviation (see Figure 3.9). We can conclude that this mapping function is expected to work properly according to the property of the desired function shown in Figure 3.4, on a wider class of images. The advantage here is that when mapping images of low noise variations, small σ, in the background (e.g. ELM images) then the function tends not to magnify that noise.

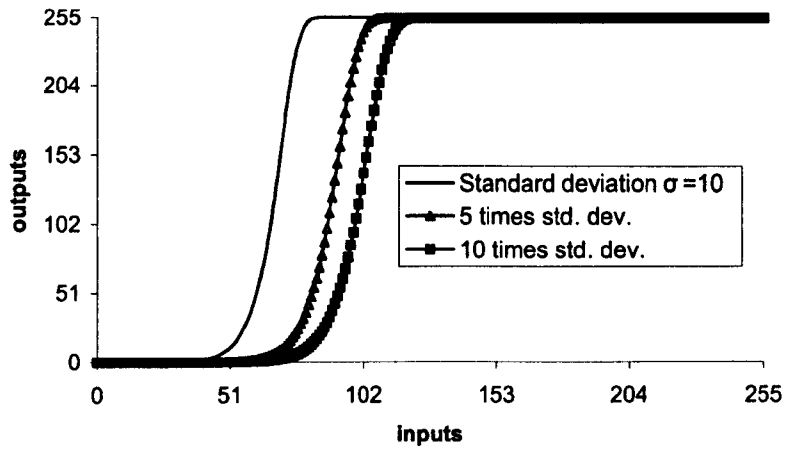


Figure 3.9: Mapping function $F(\Phi)$ at three different standard deviations: σ =10, 50, and 100.

The selection of the standard deviation (σ) of this mapping function is automatically determined according to the estimated standard deviation of the background surrounding the lesion. A lesion containing large variations in the background needs a larger standard deviation, whereas lesser variations need a smaller σ. Another appealing feature of this method is that the choice between seven different smoothing kernels (minimum of 3x3 pixels and maximum of 15x15 pixels) is dynamically selected according to the level of noise estimated from the standard deviation of background samples.

3.3.2 Isodata algorithm

The iterative selection technique for choosing a threshold was developed by Ridler and Calvard [Ridler and Calvard, 1978] and was described by Madisett et al. as an isodata algorithm [Madisetti and Williams, 1998]. It is used to obtain a threshold value for an image, whereby an optimum auto-threshold value T is chosen as the result of an iterative process. Repetitive iterations provide increasingly cleaner extractions of the object region. This value segments the image into binary regions containing the lesion and the background. Here the histogram is initially segmented into two parts using an initial threshold value of $T_0=2^{(B-1)}$, where B is the number of bits. For an 8-bit intensity image $B=8$ and $T_0=128$. The sample mean of the lesion associated with foreground pixels $(m_{f,0})$ and the sample mean of the intensities with the background pixels $(m_{b,0})$ are computed. A new generated threshold value T_k is computed as the average of these two sample means. The process is repeated, based

upon the new generated threshold, until the threshold value, which is an integer, does not change any more [Hance, et al., 1996]:

$$T_k = \left(m_{f,k-1} + m_{b,k-1}\right)/2 \qquad until \quad T_k = T_{k-1} \tag{3.9}$$

This approach is different to the thresholding technique used by Xu et al [Xu, et al., 1999], which uses the average intensity of the p% highest gradient pixels in the image to determine *T*.

3.3.3 Object outlining

Object outlining [Rasband, 2000] simply follows the logic rule that "any foreground pixel with at least one background pixel in the 3x3 neighborhood is left unchanged, otherwise it is changed to the background colour". The equivalent pseudo-code is shown in Table 1.

Table 1. Pseudo-code for determining object outline

```
{Outline object}
    p = p5              {p5 is the central pixel in a 3x3 window}
    if (p ≠ bgColor) then {bgColor is the background colour}
        if all neighbourhoods pixels of p5 ≠ bgColor then
        p = bgColor                 {p is not an Edge pixel}
    endif
```

3.4 Neural network edge detection (NNED)

A novel neural network training set, consisting of a limited number of prototype edge patterns (See Figure 3.10), is proposed here to establish a base for a simple neural network edge detection scheme. The generated set of vectors simply corresponds to standard situations that are unequivocally understood as edges or non-edges and permit a controlled distribution of the edges enclosed. Based on this scheme, chapter 4 will contain a discussion of various NN experiments with different training set sizes as an attempt to overcome this drawback.

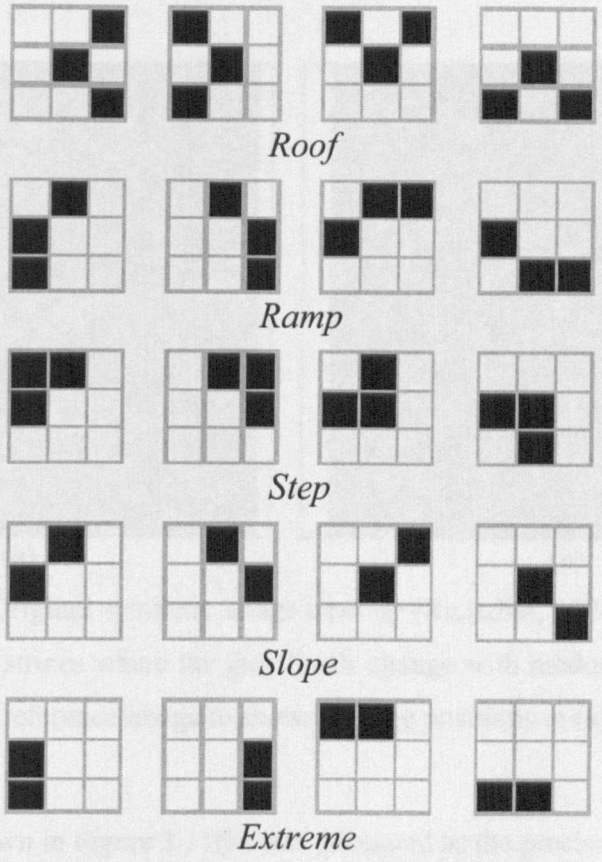
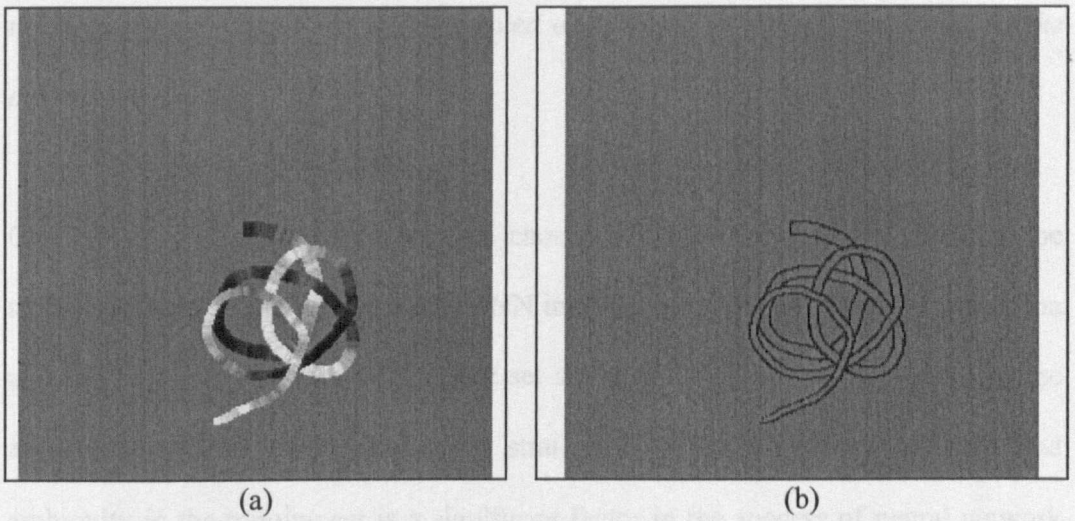


Figure 3.10: Example of five edge profiles in a 3x3 pixels grid, at four orientations.

The scheme we have used is straightforward and easier than a scheme that generates an image containing more complex edge patterns possessing variation of intensities. At the time of generation, a second type of image is also produced to show the exact edge positions (edges and non-edges patterns can be collected from this reference image). The accuracy of the reference edge map is important since it is considered as a basis for any further comparisons (e.g. comparisons between different edge maps which might be obtained from different edge detection techniques). An example of this scheme can be illustrated by the two synthetic images generated by Ramalho [Ramalho, 1996] which are shown in Figure 3.11(a) and (b).



(a) (b)

Figure 3.11: (a) Original synthetic image used by [Ramalho, 1996] which consists of random radius stripes where the grey levels change with random slopes. (b) The second generated reference image to show the edge positions in (a).

The edge map shown in Figure 3.11(b) was generated as the precise location of edges (of the original image Figure 3.11(a)) and to compare the performance of different

conventional edge approaches (e.g. Roberts [Roberts, 1965], Canny [Canny, 1986], ...etc) along with the derived neural network edge detector.

In spite of the fact that the reference edge map (Figure 3.11(b)) does not mimic all the edges present in the natural scenes, the use of this image along with two more distinct edge maps from two different conventional edge approaches (for example Roberts versus Canny) have established a basic arbitration strategy using neural networks [Ramalho, 1996]. The research here concluded that some improvement could be achieved by such strategy because *the neural network (described as an edge arbitrator) can inherit some of the edge features (e.g. sharpness, contrast, and thickness of an edge) from the arbitrated edge maps, which will finally assist the arbitration process.*

One of the advantages of the scheme chosen in this work is that it may also be suitable to analyse the behavior of the NN investigated because of the generalization and the flexibility to alter the training set structure (size, orientation, and also the amount of random noise added). The strategy of minimizing the complexity and ambiguity in the training set is a significant factor in the success of neural network recognition. This is a more important factor than minimizing the convergence error of the trained neural network [Nakashima, et al., 2001].

In this case we use a noise free training set of a limited number of prototype edge patterns with 3x3 pixels. It is important to mention here that further experiments

could be possible with larger window sizes such as 5x5 or 7x7...etc. We choose the smallest window size of 3x3 because it is sufficient to define all edge patterns encountered in this work. Various experiments have been performed using the neural network model with multi-layer perceptron architecture. The neural network is trained with the error back-propagation algorithm [Lippmann, 1987], using different sizes of training sets. When constructing and experimenting with various training sets the total number of prototype edge patterns and their redundancy were considered. The total number of prototype edge patterns used in this work is forty-two incorporating five edge profiles with the possible orientations of each edge profile and also a hundred redundant edge patterns; this problem will be considered in Chapter 4. Figure 3.10 illustrates an example of five edge profiles in 3x3 pixel grids.

The training procedure is followed until a compromise solution is achieved and the neural network is capable of detecting object boundaries. The work presented here was based on a neural network model of the three layer multi-layer perceptron architecture. An example of 9-10-7 multi-layer perceptron MLP is shown in Figure 3.12: where there are nine inputs in the input layer, ten hidden nodes (units), and seven nodes in its output layer.

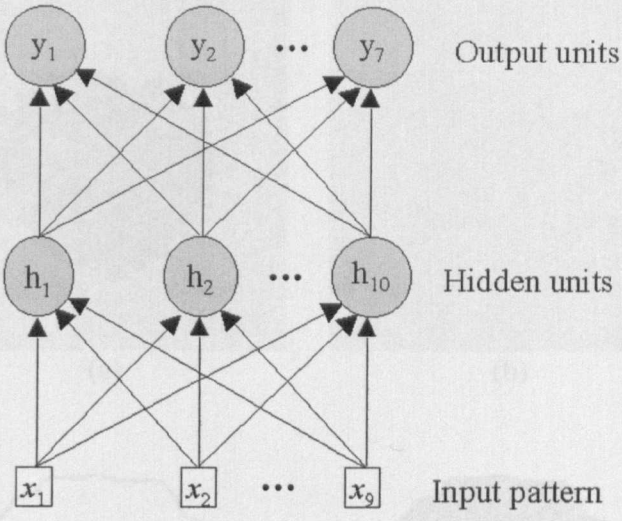
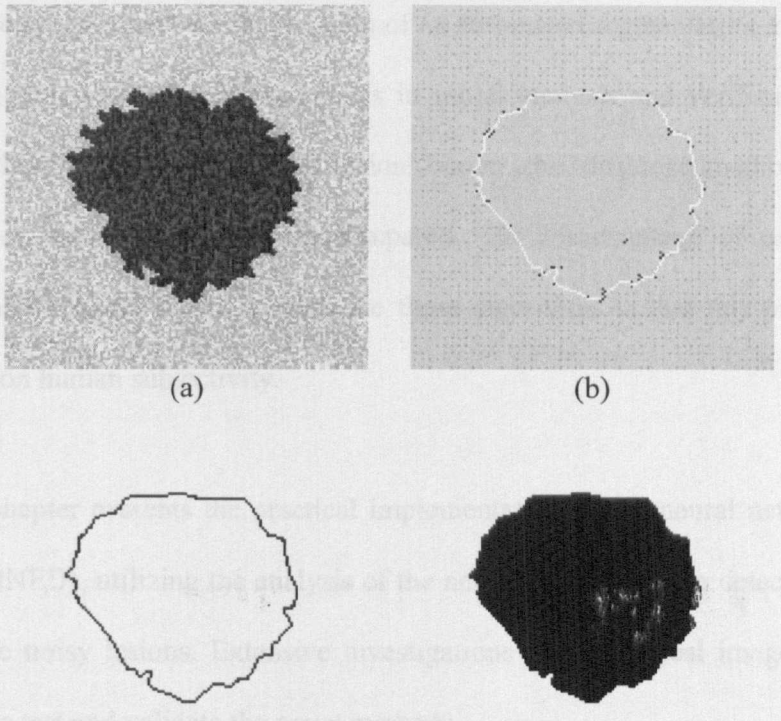


Figure 3.12: 9-10-7 multi-layer perceptron.

Successful training was achieved when the neural network's mean square error converges to the value of 0.009 with learning rate of 0.001 and momentum rate of 0.5 [Lippmann, 1987]. The neural network was found to be highly capable of recognizing edge patterns in noisy synthetic lesions and less likely to recognize intensity variations surrounding the lesion. Figure 3.13 illustrates the NNED method, when applied to a noisy synthetic lesion with SNR =12 (figure 3.13(a)). The edges found by the NNED are shown in figure 3.13(b). The discontinuities between edge patterns shown in figure 3.13(b) can be overcome using the same Gaussian curve fitting method as used in the ASS method [Xu, et al., 1999] (figure 3.13(c)). Using this border outline, a binary image (figure 3.13(d)) can be easily formed.



(a) (b)

(c) (d)

Figure 3.13: NNED method: (a) Example of input noisy lesion at SNR=12. (b) Highest intensities recognized edge patterns are the candidate points in the Gaussian curve fitting. (c) The white intensity edge points representing the curve shown in (b). (d) The binary area under the edges in (c).

3.5 Summary

The principle aim in the evaluation of a skin lesion image segmentation technique is to study the effectiveness in distinguishing a lesion from its background. In this chapter, two new approaches to implementing this type of technique are suggested: iterative segmentation (IS) and neural network edge detection (NNED). These novel techniques will be compared in Chapter 5 with the established automatic skin segmentation method (ASS). The aim is to quantitatively analyze the error in

locating the border due to the application of an automated segmentation method. The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be quantitatively and more accurately compared. The disadvantage of using expert delineations of actual lesions to evaluate these algorithms is that this procedure is dependent on human subjectivity.

The next chapter presents the practical implementation of the neural network edge detector (NNED), utilizing the analysis of the neural networks edge detector applied to synthetic noisy lesions. Extensive investigations of simple real images are also presented to test and validate the novel methods.

References

Canny J, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, no. 6, pp. 679-698, 1986.

Clark JJ, "Authenticating edges produced by zero crossing algorithms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 1, pp. 43-57, 1989.

Cotton S, Claridge E, et al., "Noninvasive skin imaging," in Information Processing in Medical Imaging, Springer-Verlag, Poultney, Vermont, 1997, pp. 501-506.

Cotton S, Claridge E, et al., "A skin image method based on a colour formation model and its application to the diagnosis of pigmented skin lesion," in Medical Image Understanding and Analysis, 1999, pp. 49-52.

Golston JE, Stoecker WV, et al., "Automatic detection of irregular borders in melanoma and other skin tumors," *Computerized Medical Imaging and Graphics*, vol. 16, no. 3, pp. 199-203, 1992.

Goshtasby A, "Geometric modelling using rational Gaussian curves and surfaces," *Computer-Aided Design*, vol. 27, no. 5, pp. 363-375, 1995.

Gray SB, "Local properties of binary images in two dimensions," *IEEE Trans. Computer*, vol. 20, no. 5, pp. 551-561, 1971.

Hance GA, Umbaugh SE, et al., "Unsupervised color image segmentation with application to skin tumor borders," *IEEE Eng. Med. Biol. Mag.,*, vol. 15, no. 1, pp. 104-111, 1996.

Hill B, Roger T, et al., "Comparative analysis of the quantization of color spaces on the basis of the CIELAB color-difference formula," *ACM Trans. Graphics*, vol. 16, no. 2, pp. 109-154, 1997.

Kasson JK and Plouffe W, "An analysis of selected computer interchange color spaces," *ACM Trans. Graphics*, vol. 11, no. 4, pp. 373-405, 1992.

Lee TK, McLean DI, et al., "Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions," *Medical image analysis*, vol. 7, no. 1, pp. 47-64, 2003.

Lippmann RP, "An Introduction to Computing with Neural Nets," *IEEE Acoustic Speech and Signal Processing Magazine*, vol. 4, no. 2, pp. 4-22, 1987.

Madisetti VK and Williams DB, *The digital signal processing handbook*: CRC Press, Boca Raton FL, 1998.

Moncrieff M, Cotton S, et al., "Spectrophotometric intracutaneous analysis: a new technique for imaging pigmented skin lesion," *British Journal of Dermatology*, vol. 146, no. 3, pp. 448-457, 2002.

Nakashima A, Hirabayashi A, et al., "Error correcting memorization learning for noisy training examples," *Neural Networks*, vol. 14, pp. 79-92, 2001.

Ramalho M, "Edge detection using neural network arbitration," PhD thesis, School of Elect. and Electron. Eng., University of Nottingham, Nott., UK, 1996.

Rasband W. (2000). Scion image for windows. e-mail: wayne@codon.nih.gov. http://www.scioncorp.com

Ridler TW and Calvard S, "Picture thresholding using an iterative selection method," *IEEE Trans. on Systems, Man and Cybernetics*, 1978.

Roberts L, *Machine perception of three dimensional solids*: MIT Press, 1965.

Rosenfeld A, "Extraction of topological information from digital images," University of Maryland, College Park, MD: Computer Science Center, Tech. Rep. TR-547, 1977.

Satta F and Claridge E, "Detection of blood deprived regions in SIAgraph images of pigmented skin lesions," in Medical Image Understanding and Analysis, 2002.

Xu L, Jackowski M, et al., "Segmentation of skin cancer images," *Image Visions Computing*, vol. 17, pp. 65-74, 1999. http://www.cs.wright.edu/people/faculty/agoshtas/paper_fig.html

CHAPTER 4: ANALYSIS OF NNED

This chapter presents a new method to analyse neural networks for edge detection so as to achieve a close insight into their internal functionality. To this purpose, a new training set, consisting of a limited number of edge patterns, is proposed to analyse the problem of neural network edge detection for both synthetic and real life images. The training set with a limited number of patterns is required in the case of edge recognition and this would keep the computational cost to a minimum. The class of images used in the testing phase of NNED is mainly synthetic lesions; this is expected to provide some sort of pixel classification; every detected pixel is either part of an edge border or not. Moreover, any success achieved in the generalised NNED will be used to establish general trends of errors for the NNED segmentation method for cluster of noisy synthetic lesions (Chapter 5). Some other selected colour real images after being simplified to intensities mapped images (Chapter 3) are also used along with some standard real images in order to assist this test. A generalization is achieved, with reduced computational time, for those training sets containing sharp edges.

4.1 Introduction

In the past 10 years, researchers have attempted to analyse the internal functionality of neural networks. The main purpose of this analysis was to arrive at a closer understanding of the internal behavior of different neural networks systems for

different applications. For example, great improvements have been made by the rule extraction method [Craven and Shavlik, 1994] in decision-making systems and other systems that can easily be expressed as a set of rules. Another example of the so-called sensitivity analysis (nonparametric statistical analysis method) [Hashem, 1992] was also used to analyse neural networks systems with relatively few inputs. However, for both methods it would be difficult to analyse most of neural network systems with so high a dimension. This is because extracting rules becomes too large to interpret by the rule extraction method, or so nonlinear for the sensitivity analysis method (perhaps part of the input space could be valid for this analysis) [van der Zwang and Slump, 2002]. A neural network analysis method was introduced and presented in earlier publications by van der Zwang [van der Zwang, et al., Sept. 2002;van der Zwang, et al., March 2002]. This method utilizes domain-specific base functions that are easy to interpret by the user and can even be used to optimize neural network systems. For example, a problem domain, in a recent publication by van der Zwang [van der Zwang and Slump, 2002], the case of edge detection - a digital image processing technique - was investigated and in particular the 2-dimensional differential operators (base functions). It was also found that it is feasible to analyse the trained neural network's edge detection in terms of gradient filter components (zero-order or low-pass, gradient, and second-order gradients; in a similar manner when analysing other digital image operators) and this will enhance the understanding of the neural net's functionality [van der Zwang and Slump, 2002].

In image segmentation, which basically considers an image as a combination of segments within which image data are more or less homogeneous [van der Zwang and Slump, 2002], neural network edge detection is frequently used. Searching for these segments is mostly done by two main approaches: (1) classification of all pixels is based on a criterion of homogeneousness, (2) detection of all pixels that are lying between the borders of homogeneous areas. In our work with the NNED, the second approach is more appropriate. In fact, the neural network edge detection approach can also be thought as a sort of pixel classification; every pixel is either part of an edge or not and hence all edges together might form different contours of different segments [van der Zwang and Slump, 2002]. In practice, not all pixels are classified correctly due to noise, etc. Therefore an extra step of edge linking is used sometimes in order to obtain closed contours.

Synthetic images were also used to support the analysis of neural network edge detection so as to gain insight into the internal functionality of this technique. Zwang et al. used the two training images shown in figure 4.1 to analyse a neural network edge detector; the neural networks were of the feed-forward error-backpropagation type. Hidden nodes, whose weights were regarded as a template, which is similar to any image filter (e.g. Kirsch or Sobel templates), and its Taylor series coefficients, were used to analyse the order of this template. Interesting results were found when some small neural networks edge detectors were trained with sharp edges shown in Figure 4.1 whilst others were trained with sharp, blurred, and noisy variants of the same images. For the first category of neural network, the Taylor series coefficients

analysis shows that most hidden units act as both a gradient filter and also have second-order gradient behaviour, but they do not have a significant low-pass (zero-order gradient) component. When analysed the second category of network's hidden units have similar gradient behaviors as the previous one, while the second-order gradient components are somewhat stronger. On the other hand, the low-pass components are notably present. It is worth mentioning that the amount of noise was not precisely specified in the method of Zwang et al. method while analyzing various neural network architectures. Moreover, various neural networks that have been analysed only consist of one single output, whereas we have analysed similar neural network architectures but with multi-outputs.

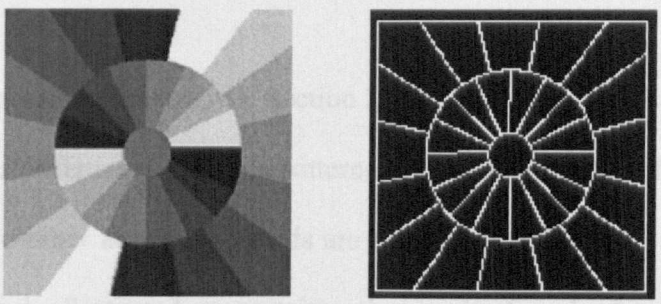


Figure 4.1: van der Zwang et al. input training image and its reference edge map.

In our approach, a standard benchmark for generating a model for the noise, modeled as Gaussian noise deviates, is considered when generating synthetic prototype edge patterns. This approach has an advantage of *flexibility* and *generality* to generate an unlimited number of training sets. The results obtained when varying the number of

noise levels are compared in order to establish the degree of effectiveness of this approach. The major disadvantage of using a vigorous noisy training set is that the trained network memorizes the noisy examples by rote [Nakashima, et al., 2001], the so-called "overfitting" phenomenon. Therefore the neural network would memorize the prototype noisy edge patterns when an image contains noise patterns similar to those included in the training set. However, the aim of using this approach is mainly to offer an abundance of preliminary test experiments for its validity (see sections 4.4.1.1 to 4.4.1.4). Moreover, experiments applied to non-noisy training sets, containing diverse number of edge patterns, are also presented (see section 4.4.1.4). The non-noisy training set could be considered as a case of a set containing noise but at a very low level of noise standard deviation (or S/N >> 1).

This chapter is organized as follows. Section 2 starts with an introduction of the tools used to generate different noisy edge patterns. The images that are formed by these patterns and also other adopted methods are described, in Section 3, to achieve the aims of this work. Section 4 summarizes the results found when evaluating the success rate of the investigated neural network edge detector. Finally, in Section 5 conclusions and suggestions for future work are considered.

4.2 Analysis Method

The noise model, which is added to the neural network training set, is discussed first:

4.2.1 Gaussian Noise

In our case, the noise is chosen to be additive [Garcya-Silvente, et al., 1997] to the prototype edge patterns. Therefore, a generated noise image $n(i, j)$ added to the true image, free-of-noise, $I(i, j)$ will yield $I_n(i, j)$ as

$$I_n(i,j) = I(i,j) + n(i,j) \tag{4.1}$$

This model of Gaussian noise deviates [Press, et al., 1989] is chosen because:

a) One can define the Gaussian noise values as normally distributed deviates with zero mean and at fixed standard deviations, σ_n as $n(i, j;\sigma_n)$. These noise values are completely uncorrelated with each other and also with the true image e.g. $I(i, j)$. Noise is considered to be additive.

b) The symmetrical nature of this distribution is a practical approximation. In contrast impulsive noise, otherwise known as spots or peak noise, randomly alters the pixels so that their values deviate considerably from the true values $I(i, j)$.

Two Gaussian deviates are generated by the algorithm introduced by Press et al. [Press, et al., 1989]. This algorithm has been developed to generate these normal deviates at the desired σ_n and in two dimensions i and j to yield finally a noise image, file, of size nxm.

4.2.2 Neural Network Training Set

In the present work, neural network training sets consist primarily of potential prototype edge patterns. Figure 4.2 shows 4 types of edges of various orientations. This NN based model may provide a

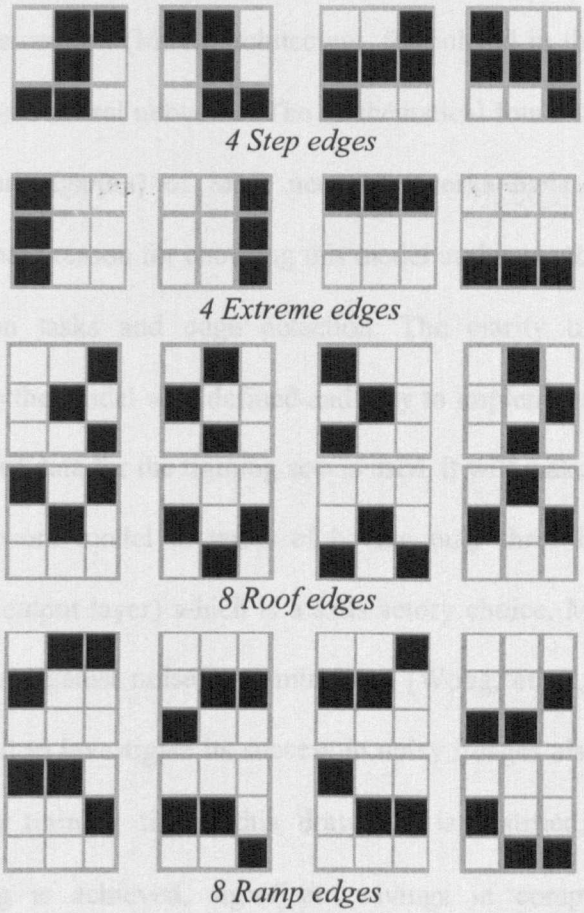


Figure 4.2: Twenty-four different prototype edge patterns in a 3×3 pixels grid representing four edge profiles.

a shortcut for finding the local maximum edge strength (e.g. Eqn. (4.6) see later), since the computational effort to calculate four edge strengths can be reduced into one step. It should be noticed that these patterns are subject to further tuning in the proceeding work until successful training is achieved.

4.2.3 Neural Network Design

The multi-layer perceptron (MLP) architecture, formulated in the 1960's, is one of the most widely used neural networks. The mathematical foundations of the training algorithm (back-propagation) of MLP neural networks have been addressed in Section 2.4. The main reason for choosing this model in this work is its suitability for pattern recognition tasks and edge detection. The clarity of its mathematical foundations makes the model well defined and easy to implement. In our application when a small size of data for the training sets is used, it will make the utilization of a simple neural network model in terms of having only three layers (input layer, hidden layer, and output layer) which is a satisfactory choice. Moreover, since this neural net is robust against noise contaminations [Wong, et al., 2000], this would make our final goal to investigate its success in noisy images also possible. In spite of relatively long training times, this drawback is justified, because when a successful training is achieved, significant savings in computational load are obtained in running the generalized neural network [Khashman, 1997]. This keeps the computational cost to a minimum for the edge detection scheme.

The work presented here is based on a neural network model of three-layer multi-layer perceptron architecture. An example of 9-10-7 MLP is shown in Fig. 4.3: where there are nine inputs in the input layer, ten hidden nodes (units), and seven nodes in its output layer.

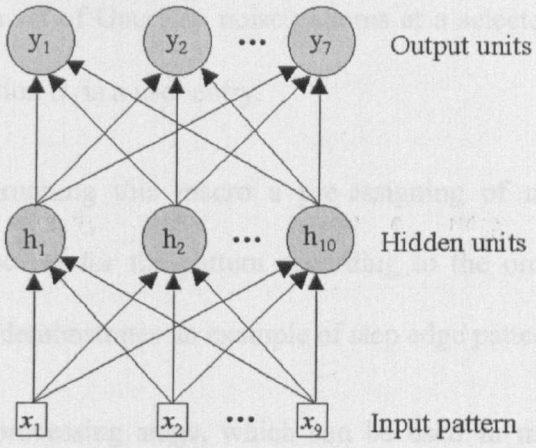


Figure 4.3: 9-10-7 multi-layer perceptron.

4.3 Procedure

In this Section the image processing software used to generate various prototype edge patterns is introduced. Examples illustrating the technique used in encoding edge patterns and then decoding their outputs from the neural network are also considered. Firstly, the software used in generating various prototype edge patterns is described.

4.3.1 Generating prototype edge patterns

Scion Image software [Rasband, 2000] was used as a developing environment to achieve this goal. Firstly the flowchart of structure control of SNEPG: a **S**et of **N**oisy **E**dge **P**attern **G**enerator software macros is demonstrated in Figure 4.4.

Macro A: generate a set of Gaussian noise patterns at a selected level of noise; the noise standard deviation σ_n is a user entry.

Macro B: before running this macro a pre-assigning of nine values of pixel amplitude need to be set for the pattern according to the order shown in Figure 4.5(a). Figure 4.5(b) demonstrates an example of step edge pattern.

Macro C: is a pre-processing stage, which can be used in most neural networks applications [Tarassenko, 1998]. In the present work, data is normalized between two values 0.0 and 1.0 after being shifted to zero (a possibility of negative values are due to noise data so that a minimum value within the set of data needs to be found). If $E_n(i, j)$ is a set of edge patterns after noise is added to it, then $E_{sf}(i, j)$ is the shifted signal given as follows

$$E_{sf}(i,j) = \left|\min(E_n(i,j))\right| + E_n(i,j) \tag{4.2}$$

Macro D: is an essential part in encoding data by which the outputs of this macro form the target outputs of the neural network. Two entries are required in this macro:

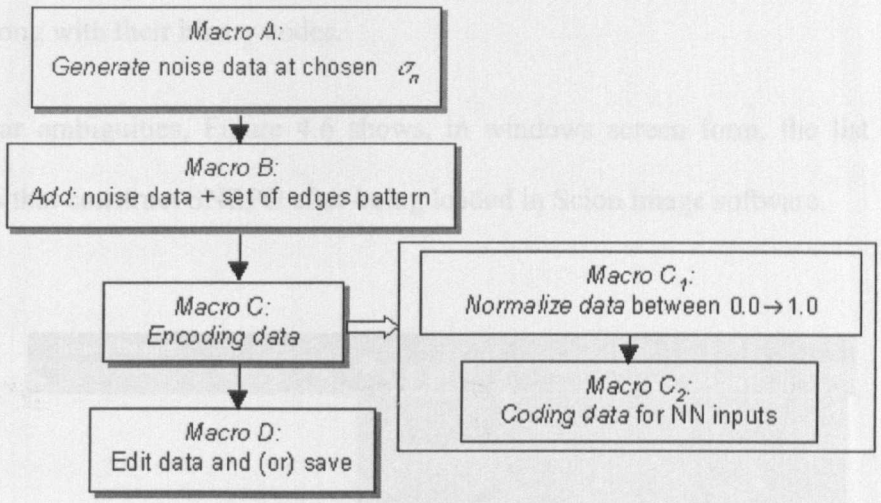


Figure 4.4: Flowchart of structure control of SNEPG.

A	*B*	*C*
D	*E*	*F*
G	*H*	*I*

(a)

A	*B=1*	*C=1*
D	*E=1*	*F*
G=1	*H=1*	*I*

(b)

Figure 4.5: (a) Notation used. (b) Pixel values representing step edge pattern.

(i) number of bits used to encode data which is the same as the number of neural net's outputs, and (ii) code to be generated for the whole set of edge patterns; for simplicity entries are designed in decimal. However, the macro will output the same data along with their binary codes.

To clear ambiguities, Figure 4.6 shows, in windows screen form, the list of five macros that construct SNEPG after being loaded in Scion image software.

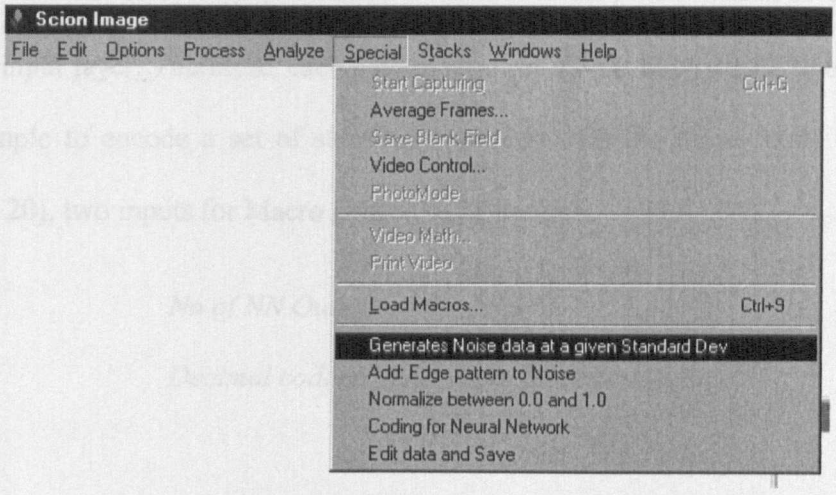


Figure 4.6: SNEPG consists of five macros.

4.3.2 NN Inputs Encoding

As discussed in the previous section, SNEPG macros were used in the pre-processing stage of the neural network. The entries of the encoding macro are defined in this section (Macro D). At the early stages of neural network experiments, the total number of prototype patterns considered was set at 24 (e.g. Figure 4.2). The whole set of the edge patterns will be corrupted by a few noise levels. We have also chosen

five noise levels from low to high as a maximum number of levels, since adding too many levels of noise would cause the neural network to overfit in other words the trained network memorizes the noisy examples by rote [Nakashima, et al., 2001]. Therefore, adjusting the value of standard deviation of the noise generated or the S/N ratios for the prototype edge pattern set, is the subject of this chapter (experimental results are summarized in Section 4.4) until we reach to the generalized solution for NNED. To estimate the total number of neural network outputs, we have to consider that 120 different noisy edge patterns (24×5 *noise levels*) are required in the neural network input layer. Therefore, each pattern requires a 7-bit binary code to encode it. For example to encode a set of step edge patterns with the binary code 0010100 (decimal 20), two inputs for Macro D need to be defined:

No of NN Outputs = 7

Decimal code to generate in binary = 20.0

It is interesting to mention here that to *estimate* the total number of NNED outputs, we have analysed the reference edge map (Figure 4.1) that was suggested by van der Zwang [van der Zwang and Slump, 2002]. After scanning that image, it was found that the total number of prototype edge patterns in a 3×3 window is 70 patterns. For example, more ramp patterns are contained between these scanned patterns; e.g. Figure 4.7 shows an example of another eight possible ramp patterns that are drawn in a vertical direction. Even if we also accept that total of 70 patterns, to train the

NNED in some of the possible directions and to aid our *estimation*, then we also need 7-bits in the neural network output layer to encode them.

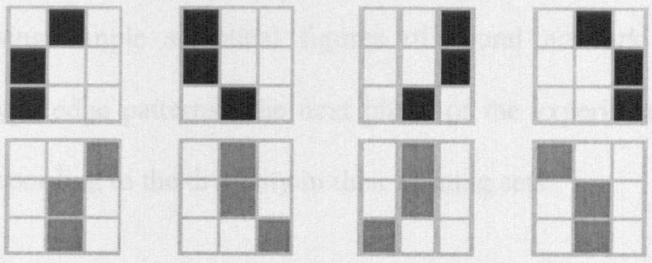


Figure 4.7: 8 possible ramp edge patterns in a vertical direction.

4.3.3 NN Outputs Decoding

After a successful training of the neural network with various noisy edge patterns, a testing phase is carried out and a decoding procedure is required to decode the winner edges from the neural network outputs. An output code which matches the following criterion is considered as a winner edge:

If $Output \in \{EdgePatternCodes\}$
$\quad winner\ edge = 1$
else
$\quad winner\ edge = 0$

In the next Section, various test experiments are performed on the neural network using the above encoding/decoding processes to test the capability of this model to locate edge maps.

4.4 Results

In this section, the experimental results for the neural network edge detection are considered. Firstly, the degree of success of the suggested neural network training set is estimated using simple statistical figures of neural network recognition for different prototype edge patterns. The next phase of the experiments is performed and presented according to the diversity in their training sets.

4.4.1 Test experiments

In order to test our analysis method, we trained several NNs for edge detection. The NNs are of a feed-forward error-back propagation type, with 3x3 inputs, 4 to 70 units in the single hidden layer, and 7 outputs. All units used sigmoid activation functions. Initial experiments were performed using the following technique in generating additive noise to various prototype edge patterns: noise is generated at five levels according to the noise signal itself i.e. noise generated at five standard deviation values: σ_N = 1.0, 10, 30, 50, and 75. Visual inspection was also used in making the decision as to how these levels are distinct from each other.

Table 4.1 describes the contents of two initial training sets and their total sizes. It should be noted that total size is calculated according to the following equation:

$$Total\ noisy\ patterns = P \times R \times N \tag{4.3}$$

$$Total\ size = P \times R \times N \times W \tag{4.4}$$

$$Total\ size(Kbytes) = P \times R \times N \times W/1024 \quad (4.5)$$

where

P = No. of patterns,

R = repetition of patterns,

N = Noise levels,

W = Window size

Since each edge pattern occupies a 3x3 window then W here is calculated as 9 bits. Repetition here means that the same prototype edge pattern is repeated R times. The aim of this is to assist in providing enough information to the input layers of a neural network; i.e. the subset of input data should be chosen so that is adequately represents the input set of a problem domain. This case is similar to "overfitting" problem; if too few input/output pairs are chosen to train the NN, then the network will "memorize" those examples and the network will not be able to generalize to new i/o pairs that it has not been explicitly trained to recognize [Ilardi, 1999].

Table 4.1: Structure of training sets. *R=50 times*

Training set	No. of patterns	Contents				Total noisy patterns	Total size (Kbytes)
		step	roof	ramp	Ext-reme		
1	20	4	8	4	4	5,000	44
2	16	4	8	4	-	4,000	35

Neural network results that are found when applied, using the first training set, to the input layer for fourteen different neural network experiments as listed in Table 4.2,

show that the neural network was unable to converge. However, when using the second training set, where the number of noisy patterns is reduced by 16.7% then the neural network is able to converge as shown in Table 4.3 below. However, in these experiments, when the neural network adjusts its weights values at each iteration in order to minimize the error level, the convergence throughout the training phase was not stable. Moreover, when testing the ability of these NNs to detect edges using standard real images and synthetic images then no edges were detected. It should be noted that real images mostly support a general test of the NN output image, where the evaluation of the effectiveness of the NN can be made visually. On the other hand, applying the NN to synthetic images could easily assist an accurate visual arbitration between correct and false localizations of edges in the recognised output image.

Table 4.2: NN results for the first training set

HN	Learning rate	Momentum rate	Final Error F.E. = 0.02	Iterations	C or NC
70	0.4	0.5	0.0260	5000	NC
30	0.08	0.3	0.0230	4000	NC
31	0.08	0.3	0.0207	4000	NC
32	0.08	0.3	0.0206	5000	NC
35	0.08	0.3	0.0213	5000	NC
20	0.4	0.5	0.0230	3000	NC
20	0.008	0.3	0.0325	5000	NC
20	0.008	0.5	0.0347	7000	NC
20	0.008	0.4	0.0342	7000	NC
32	0.08	0.5	0.0247	5000	NC
33	0.1	0.3	0.0218	2000	NC
33	0.01	0.3	0.0274	2000	NC
33	0.01	0.4	0.0245	3000	NC
33	0.009	0.5	0.0225	7000	NC

HN: Hidden Nodes C: Convergent NC: Not Convergent

Table 4.3: NN results for the second training set

HN	Learning rate	Momentum rate	Final Error F.E. = 0.02	Iterations	C or NC
70	0.1	0.3	0.0139	3500	NC
70	0.08	0.3	0.0139	10000	NC
70	0.5	0.5	0.009	3500	NC
9	0.005	0.2	0.035274	1527	C
10	0.005	0.2	0.035256	2000	NC
15	0.005	0.2	0.034860	1500	NC

HN: Hidden Nodes C: Convergent NC: Not Convergent

Two conclusions can be made from these experiments: (i) the neural network is unable to detect edges when the training sets contain a significant amount of noise (σ_N = 50 and 75), (ii) even with the reduction in the total number of noisy patterns, recognition only occurs when test images include simple prototype patterns (similar to those contained in the original training sets).

The failure of this neural network was analysed and a possible solution was suggested using two approaches:

1) Reduction in the total number of training set edge patterns, which may be achieved by reducing the total number of levels of noise added.

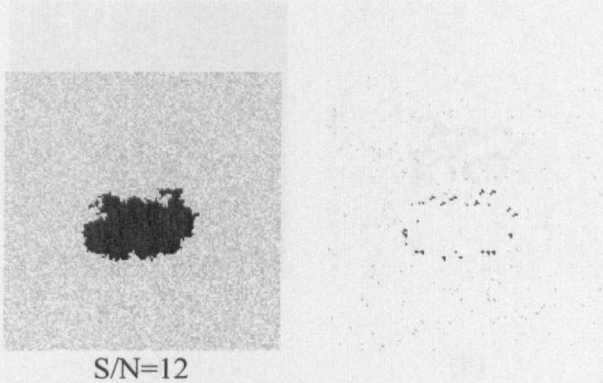
2) Applying mathematical analysis to determine the S/N ratio for each set of noisy edge patterns in the training set, to make sure that an exact amount of additive noise is the same in each pattern constructing that set.

First approach - More experiments are applied using the first approach. Similar training sets (similar to the structure shown in the first row of Table 4.1) are also generated after omitting two high noise levels (levels at noise standard deviation σ_N = 50 and 75). This will also decrease the total size of the training sets by 40%. The neural networks with $n \times n$ nodes in the single hidden layer were tested, and the first successful small neural network with 16 hidden nodes (n is equal to 4) was constructed; other NN parameters are listed in Table 4.4.

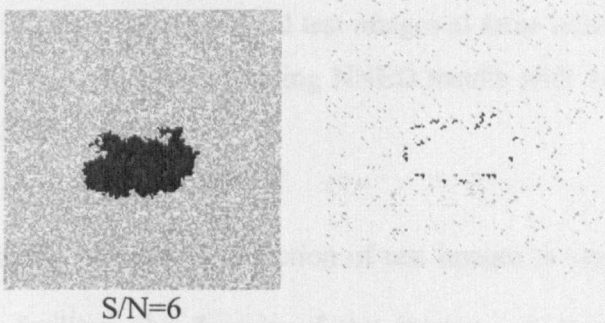
Table 4.4: Neural network parameters.

HN	Learning rate	Momentum rate	Convergent error	Iterations	C or NC
16	0.001	0.5	0.033999	633	C

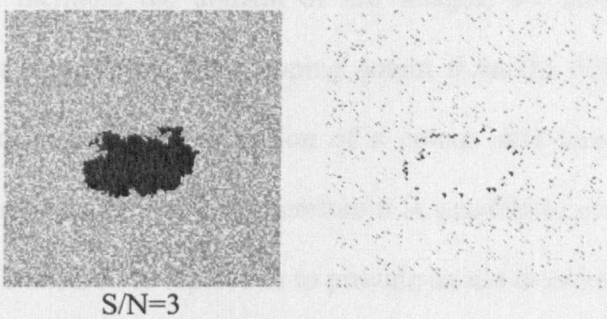
In general, larger networks (with larger hidden nodes) show different behaviours. Some functioned as noise detectors only and others detected a few edge patterns in incorrect locations. Synthetic noisy images, see Figure 4.8, give an indication of its behaviour in noisy environment. At lower S/N the NNED tends to detect fewer edge patterns. It is important to mention that the NNED outputs shown in these Figures are only representing the decoded edge patterns that are expected to represent edge positions in the corresponding original input images. The white intensities represent null detection of the NNED. Figure 4.8 compares the outputs of NNED for four noisy synthetic images with S/N=12,6,3, and 1. These images could easily provide a good test whilst experimenting with the NNED. By simple visual comparisons of the NNED results shown in Figure 4.8, we can make a decision that the NNED with 16 hidden nodes would fail to detect sharp edges when the S/N ratio is lower than 1.



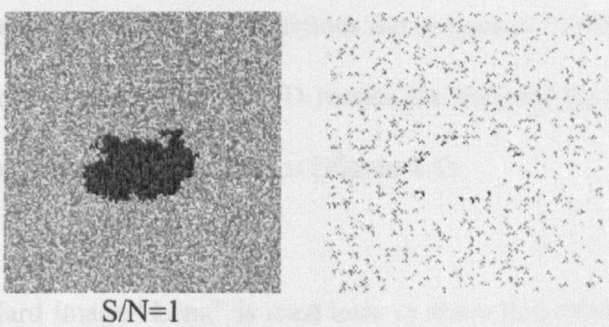
S/N=12



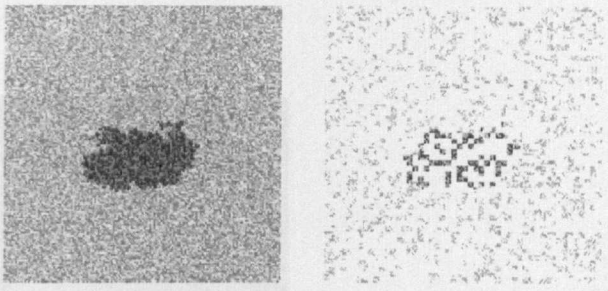
S/N=6



S/N=3



S/N=1



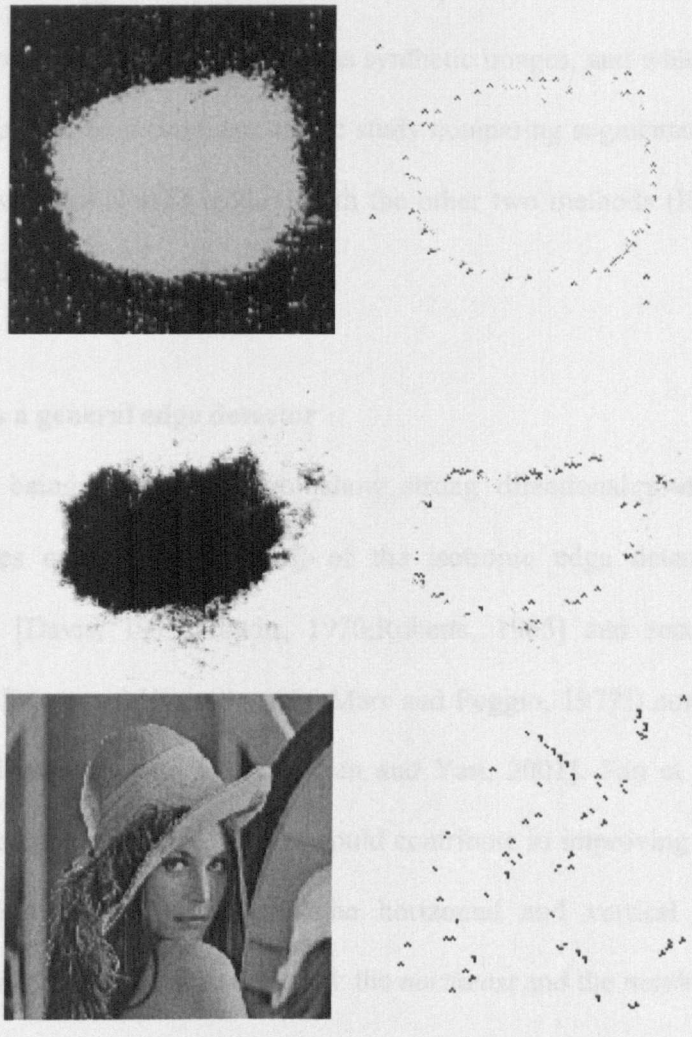
S/N=0.5

(a) (b)

Figure 4.8: (a) Original test images at three noise levels, (b) Corresponding NNED results with 16 hidden nodes.

It is worth mentioning that careful selection of test images is very important in the design stage. To facilitate the domain of test images, we have chosen the real intensities images representing the mapping output from the ASS method (Figure 3.4(c) in Chapter 3). Since segmentation of a colour skin lesion needs accurate processing to localize lesion borders, therefore it is sensible to check the possibility of using the NNED method to replace or to provide an aid in accurate localization of the optimal lesion border. Figure 4.9 compares the outputs of NNED, with 16 hidden nodes, for three real images. The NNED results for the real lesion are compatible with the results found for synthetic lesions (Figure 4.8).

The general standard image "Lena" is used here to show how NNs work as an edge detector (edges are enhanced to the highest intensities for good viewing). Since the intention here is not to enhance the ability of the NNED to work for general scene images, instead we are trying to search for a trained NNED that could achieve both



(a) (b)

Figure 4.9: (a) Recalling images from Figure 3.4(c), mapped image of colour lesion of Figure 3.1, and a standard image 'Lena'. (b) Corresponding NNED results with 16 hidden nodes.

the generalization and which also will support an accurate delineation of lesion boundaries in the suggested NNED method (Chapter 3). The successful training set that will support generalization for various synthetic images, and which is also more robust to noise, will be a candidate in the study comparing segmentation errors that are obtained with the NNED method with the other two methods (IS and ASS) as explained in Chapter 5.

4.4.1.1 Test as a general edge detector

Since human beings usually do not show strong directional preferences when detecting edges or boundaries, most of the isotropic edge detectors (gradient operators e.g. [Davis, 1975;Prewitt, 1970;Roberts, 1965] and second derivative operators e.g. [Marr and Hildreth, 1980;Marr and Poggio, 1979]) consider only the horizontal and vertical edge patterns [Fan and Yau, 2001]. Fan et al. found that including more potential edge patterns would contribute in improving their isotropic colour-edge detector. In addition to the horizontal and vertical representative patterns, two more patterns were included: the *northeast* and the *northwest* diagonals (See figure 4.10). The maximum of the four edge strength patterns are applied to a colour image to calculate the gradients in each colour component (YUV colour space was chosen [Fan and Yau, 2001]). For example the local maximum edge strength of a central pixel (i,j), $MOE_Y(i,j)$ could be defined for the luminance component (Y) as:

$$MOE_Y(i,j) = \max\{HOE_Y(i,j), VOE_Y(i,j), NOE_Y(i,j), SOE_Y(i,j)\} \qquad (4.6)$$

1	2	1
0	0	0
-1	-2	-1

Horizontal
$HOE(i, j)$

1	0	-1
2	0	-2
1	0	-1

Vertical
$VOE(i, j)$

2	1	0
1	0	-1
0	-1	-2

Northeast
$NOE(i, j)$

0	1	2
-1	0	1
-2	-1	0

Northwest
$SOE(i, j)$

Figure 4.10: Four weight-coefficients for calculating edge strength of potential edge patterns for central pixel at (i, j). For example, the edge strengths $HOE(i, j)$ are calculated as a weighted sum of the six pixel values in neighborhood of central pixel at (i, j).

Figure 4.11 demonstrates the major steps of Fan et al. colour edge detector. The first step calculates the gradients of each colour components, i.e. $MOE_Y(i, j)$, $MOE_U(i, j)$, $MOE_V(i, j)$ which are denoted in Figure 4.11 as *Gradient_Y*, *Gradient_U*, *Gradient_V*, respectively. Then, thresholding each gradient component according to an optimal threshold [Fan and Yau, 2001] to find three corresponding edge components Y_E, U_E, V_E which are denoted in Figure 4.11 as *Edges_Y*, *Edges_U*, *Edges_V*, respectively. Edge results are finally integrated through the fusion rule between every pixel in these three edges, as

$$E = \begin{cases} 1, \text{edge pixel}, & \text{if } (Y_E = 1) \cup (U_E = 1) \cup (V_E = 1) \\ 0, \text{nonedge pixel}, & \text{otherwise} \end{cases} \tag{4.7}$$

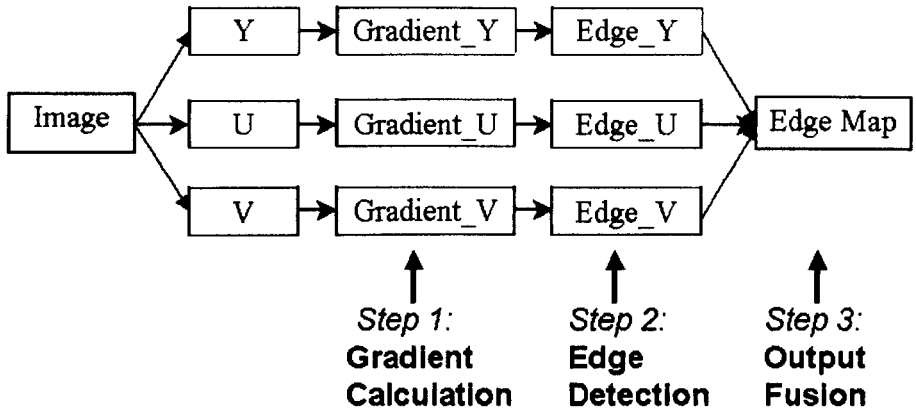
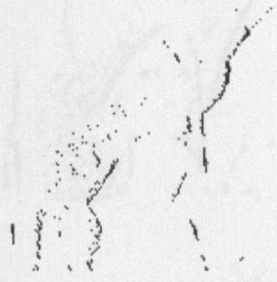


Figure 4.11: Major steps in Fan et al. edge detector.

To study the behavior of each output of the neural network, we can simply consider the weights of a neural network edge detector's hidden nodes as a template of any image filter template [van der Zwaag, et al., Sept. 2002]. The internal behavior of the neural network outputs with 16 hidden nodes and 7 outputs, which is tested in section 4.4.1, will be analysed here. Each of the seven templates of size 4×4 will be convolved with a test image to demonstrate the general behavior of each hidden node. Figure 4.12 shows an example outputs produced from seven templates. It is clearly seen from this figure that the NNED has low-pass or averaging behavior (nodes 1 and 7), which makes the network less sensitive to noise and improves its edge detection ability. We also produce for each output a corresponding example of edge strength after thresholding each output at a maximum threshold. These can also check how much details are still contained in these outputs. A fusion between outputs of edge strengths will contain so many details about the image background such that it will be difficult to make any sense of the final edge map (See example of fusion between outputs 5 and 6 in Figure 4.12).



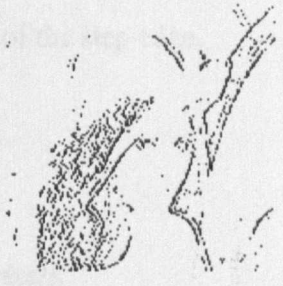
Output 1

Output 2

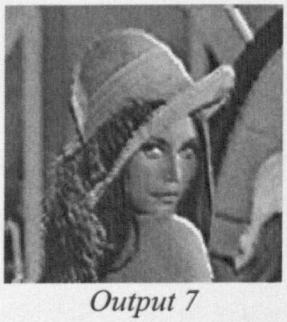
Output 3

Output 4

Output 5



Output 6



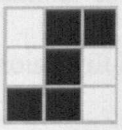
Output 7 *Edge Fusion*

Figure 4.12: NNED outputs with 16 hidden nodes. Outputs number 2 to 6 are thresholded to verify sharp edges. Output number 7 shows blurred images of "Lena". Example of edge fusion between outputs 5, and 6.

4.4.1.2 S/N analysis for prototype edge patterns

As an example, when fifty iterations of noisy step edge patterns in a 3x3 window (Figure 4.13), are generated then noise standard deviation $\sigma_N = 0.3$ is required to

have $S/N = 6$. This can be shown by simple substituting of the values of S/*N*, *N*, and S_i in equation 4.8 and solving for σ_N.



$S_{ij} = \{0,1,1,0,1,0,1,1,0\}$

Black represents 1

White represents 0

Figure 4.13: Prototype pattern of the step edge.

$$S/N = \frac{1}{N\sigma_N^2} \sum_{j=1}^{J} \sum_{i=1}^{I} S_{ij}^2 \qquad (4.8)$$

where N = *total number of pixels,*

S_{ij} = *signal pixel value,* i = *pixel index,*

j = *pattern index,*

σ_N = *standard deviation of noise*

$$\sigma_N^2 = \frac{1}{N}\frac{1}{S/N}\sum_{j=1}^{J}\sum_{i=1}^{I} S_{ij}^2 = \frac{1}{50\times 9}\frac{1}{6}\sum_{j=1}^{50}\sum_{i=1}^{9} S_{ij}^2 = 0.09$$

and

$\sigma_N = 0.3$

According to Eqn 4.8, we have calculated S/N for 50 step edge patterns (R=50) corrupted by five level of noise at σ_N = 1.0, 10, 30, 50, and 75. It is found that the corresponding calculated S/N ratios are S/N= 0.555, 0.005, 0.001, 0.0002, and 0.0001, respectively. Experimentally, we have found that the last two noise levels (at

σ_N = 50 and 75) would make the neural network fail to detect edges. The neural network would memorize the prototype noisy edge patterns when an image contains very noisy patterns similar to those included in the training set. The strategy to reduce the effects of noise will be considered in the following section.

4.4.1.3 NNED training set according to S/N

Another S/N criteria was implemented as follows: three S/N ratios were chosen: (i) no noise, (ii) 12, and (iii) 6. For example this would correspond to relatively low standard deviations of σ_N = 0, 0.215, and 0.3043, respectively. Therefore, this reduction in noise could satisfy the two criteria which are mentioned above (section 4.4.1). The above design modifications are applied to the next phase of neural network training sets and considered in the following experiments.

Similarly, amongst numerous tested training sets, an example of a successful training set structure is presented in Tables 4.5. Two neural network parameters are listed in Table 4.6 below.

Table 4.5: Structure of the training set. *R=50 patterns*

No. of patterns	Contents				Total noisy patterns	Total size (Kbytes)
	step	roof	ramp	Ext-reme		
20	4	4	8	4	3,000	26

Table 4.6: Final neural network parameters

HN	Learning rate	Momentum rate	Convergent error	Iterations	C or NC
9	0.001	0.5	0.025000	5000	NC
16	0.001	0.5	0.0250	4917	C

It is clearly seen from Figures 4.14 and 4.15 that this NNED has a very poor performance. If we compare its output results with the NNED outputs shown in Figures 4.8 and 4.9, we will find that the NN (Figures 4.14 and 4.15) is simply detecting most of the image patterns (edges and non-edges). From the above experiments using different NN shown in Table 4.6, it is found that the larger the NN (e.g. 16 hidden nodes), the more variety in behaviour among the neural units (see Figure 4.16). This is simply because either the NNED (Figures 4.14 and 4.15) could not differentiate clearly between the learned patterns or it memorizes some of the noisy patterns. Moreover, a NN with a small number of hidden nodes (e.g. between hidden nodes 15 to 10) could randomly detect very few patterns in random positions in a tested image; this case is due to the instability of NN when converging to the required mean square error. Therefore, omitting these patterns with low noise from the training sets may produce sharp edge patterns only at the boundary region.

Similarly we have analysed the behavior of the neural network with 16 hidden nodes using the Lena image. Figure 4.16 shows an example output produced from seven templates. It is clearly seen from this figure that the seven templates show no low pass behavior, whereas the outputs of Figure 4.12 presented in section 4.4.1.1 show some low-pass behaviour (e.g. Output 7). This would make the network (Figures 4.13 to 4.15) more sensitive to noise and lessen its ability to detect edges. The more noise detected by this neural network (outputs 4 and 5 in Figure 4.16) would also make it harder for the NN to reduce its effects.

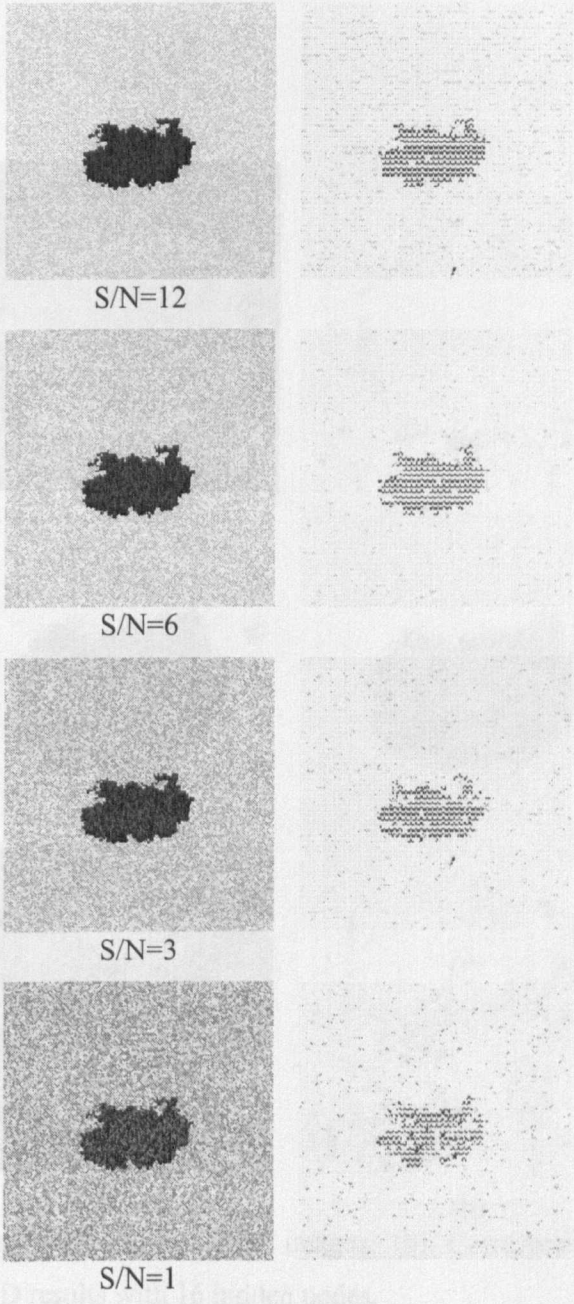
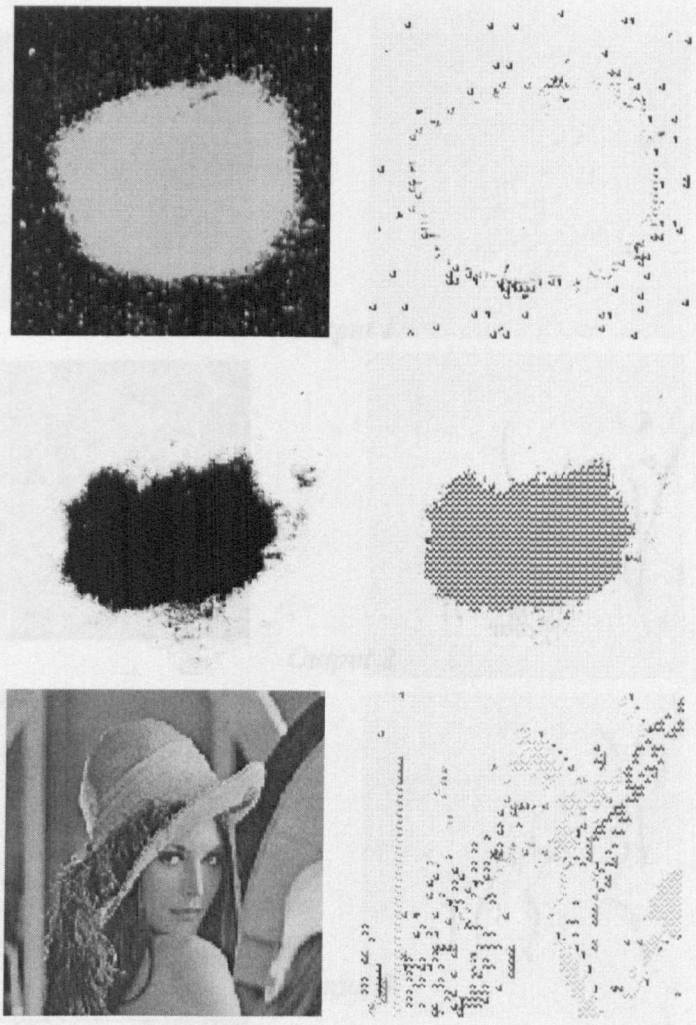


Figure 4.14: (a) Original test images at five S/Ns ratios, (b) Corresponding NNED results with 16 hidden nodes.

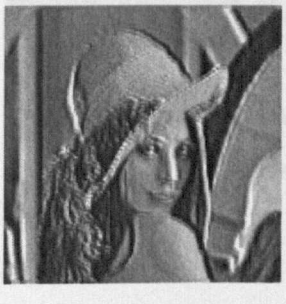


(a) (b)

Figure 4.15: (a) Original images. (b) Corresponding NNED results with 16 hidden nodes.

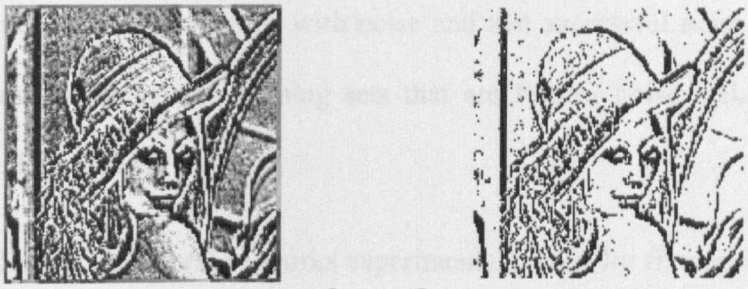
Output 1

Output 2

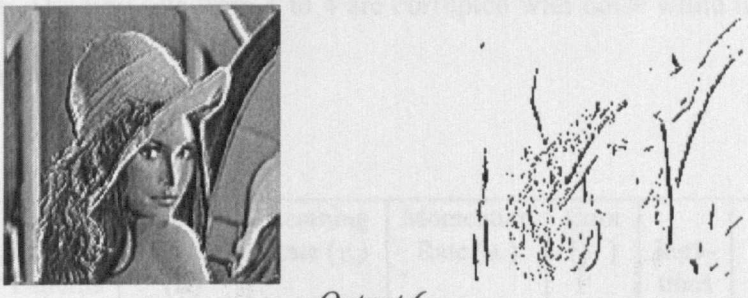


Output 3

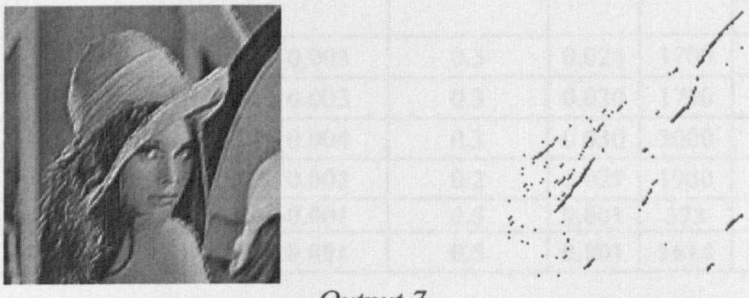
Output 4



Output 5



Output 6



Output 7

Figure 4.16: NNED outputs with 16 hidden nodes. Outputs number 1 to 7 are thresholded to verify sharp edges.

In the following sub section we will present the last experiments that are applied to the noise-free training sets.

4.4.1.4 NNED for noise-free training set

Here neural networks are trained with training sets containing sharp edges only (noise free edge patterns). Table 4.7 compare the errors between four different sizes

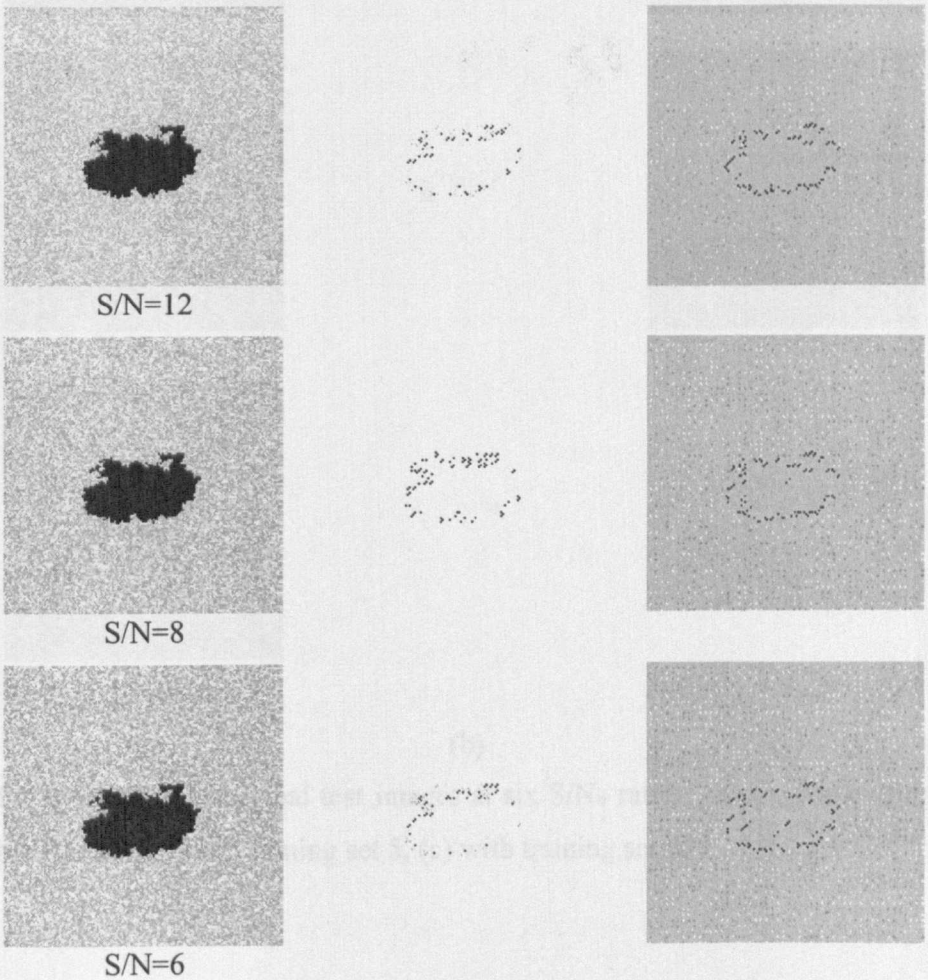
of training sets that are corrupted with noise and two successful noise free training sets. It is also seen that the training sets that are free of noise make the NNED converge with the lowest error.

Table 4.7: Successful neural networks experimental results for five different sizes of training sets. Training sets from 1 to 4 are corrupted with noise while training sets 5 and 6 are free of noise.

Training Set	Number of Edge Patterns	Hidden Nodes (H)	Learning Rate (η)	Momentum Rate (α)	Error (ε)	Itera-tions	NNED recognize more	
							noise	edges
1	40	10	0.003	0.3	0.024	1700	√	√
2	36	10	0.003	0.3	0.020	1700	√	√
3	48	10	0.004	0.3	0.030	3000	√	√
4	60	9	0.003	0.2	0.027	1980	√	√
5	**20**	**9**	**0.001**	**0.5**	**0.001**	**373**		√
6	**42**	**10**	**0.001**	**0.5**	**0.001**	**1614**		√

A remarkable improvement was shown in the neural network recognition when the total number of noise free prototype edge patterns is forty-two incorporating five-edge profiles (Figure 4.2) with the possible orientations of each edge profile and *R* was set at 150 (repetition of patterns). The aim of increasing *R* in a noise-free training set is to provide sufficient patterns to the input layer of NN while the training phase. Figures 4.17 and 4.18 shows sample results of a neural network edge detector, when applied to synthetic noisy regions, and real images. This is one of numerous experimental results that are applied to NNs when trained with different

training sets that are *free of noise* (example result is listed in rows 5 and 6 of Table 4.7). The six noisy images shown in Figure 4.17 are used as a test images along with two other similar groups of noisy images of regular and very regular lesions (Figure 3.2). Therefore, the total of 18 noisy images are used to test each training set in Table 4.7. The criterion used here is that the candidate training set is the one that makes the NNED able to detect close edge patterns, which then support accurate curve fitting.



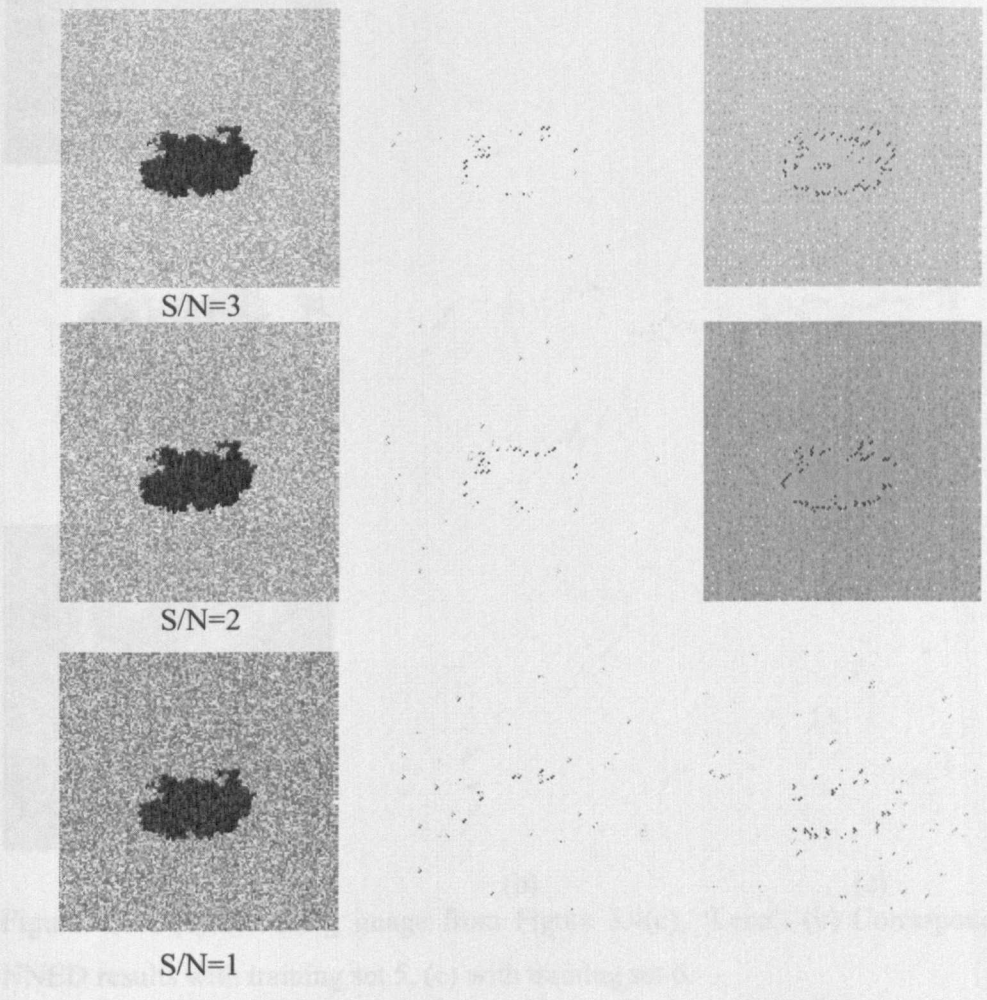
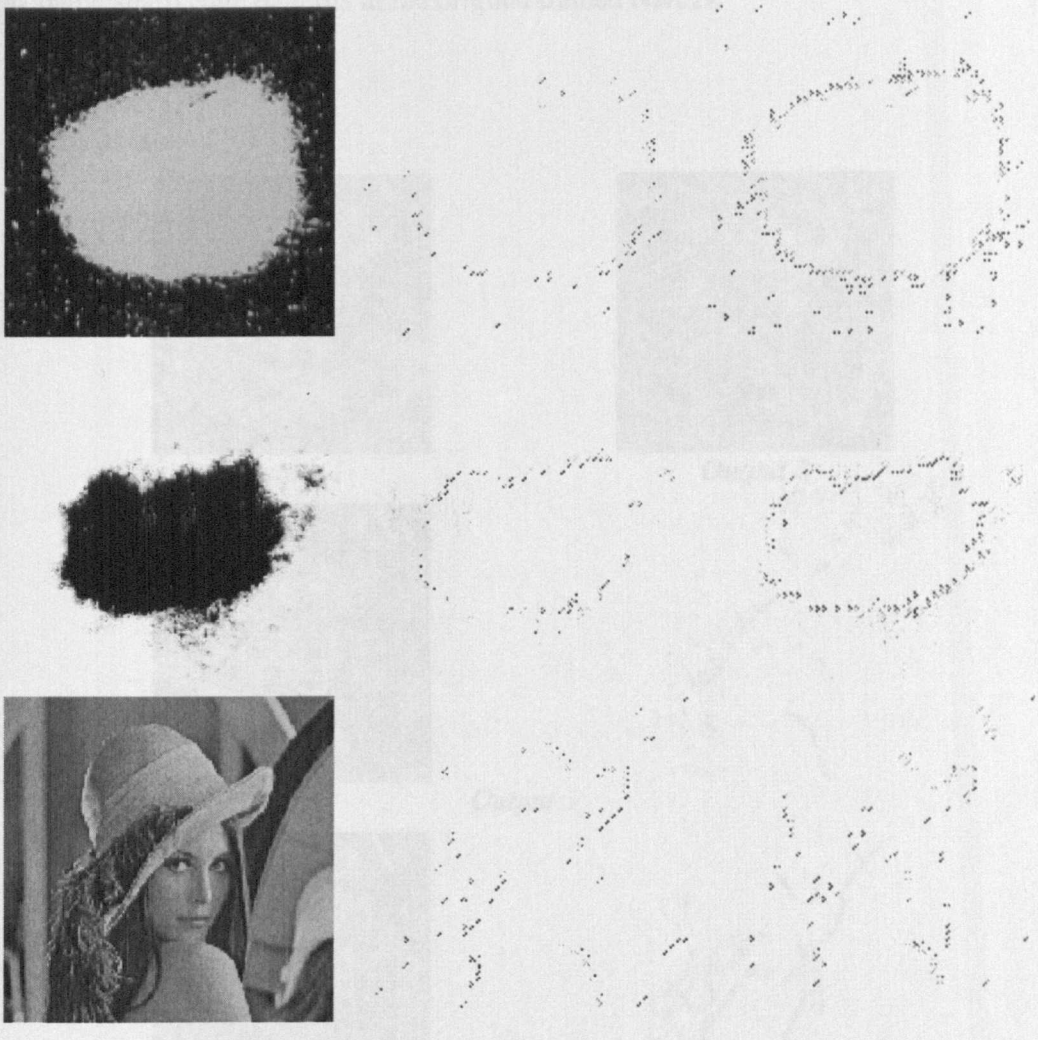


Figure 4.17: (a) Original test images at six S/Ns ratios, (b) Corresponding NNED results with training set 5, (c) with training set 6.



(a) (b) (c)

Figure 4.18: (a) Recalling image from Figure 3.4(c), 'Lena'. (b) Corresponding NNED results with training set 5, (c) with training set 6.

Low-pass or averaging behaviour makes the network less sensitive to noise and improves the edge detection ability. As shown in Figure 4.20, a fusion between three outputs of Figure 4.19 (Outputs 3,6, and 7) is compared with the output edges of the Sobel edge detector. The number of edges produced by the NNED tends to be less

than the gradients detected by Sobel filter, this is possibly due to the lack of more probable sharp edge patterns in the original trained NNED.

Output 1

Output 2

Output 3

Output 4

Output 5

Output 6

Output 7

Figure 4.19: NNED outputs with 16 hidden nodes. Outputs number 3 to 7 are thresholded to verify sharp edges. Outputs number 1 and 2 show smooth images of "Lena".

(a) (b) (c)

Figure 4.20: (a)-(b) Sobel edge detector and edge strength at a maximum threshold respectively. (c) Fusion of NNED outputs 3,4, and 7 shown in Figure 4.19.

4.5 Conclusion

In this chapter a new technique for generating a neural network training set is considered for edge detection. A limited number of noisy edge patterns, is proposed to analyse the capability of neural networks edge detection for both synthetic images and real images. Various methods were applied to reduce the neural network noisy training set; e.g. variation of training set size. However, the use of the *S/N* ratio, rather than the standard deviation of the noise, in characterising the noisy patterns in the training set was one of the successful methods which have been considered in this study to modify the technique of generating additive noise to the training set. Moreover, three levels of noise were chosen instead of five: (i) to reduce the total number of patterns (ii) to simplify the training set and (iii) to keep the computational cost of the neural network edge detector to a minimum. When constructing and experimenting on NNs with various noise free training sets, the total number of prototype edge patterns and their redundancy are considered. A remarkable improvement was shown in the neural network recognition when the total number of noise free prototype edge patterns is forty-two incorporating five edge profiles with the possible orientations of each edge profile.

Although the above only gives some analysis results for the units in the hidden units, it should be clear that a characterization of the neural network as a whole could also be derived from these results. The weights between a hidden units and the output unit represents the importance of the hidden unit's edge detection outcome. Some larger NNs have also been trained and analysed. The larger the network, the more variety in

behaviour among the neural units. In a few cases, certain units showed very strong high-order behaviour, indicating that those units functioned as noise detectors only.

Further work, using this analysis method, can be reached by analysing the behavior of different neural network architectures; e.g. RBF (Radial Basis Function) Networks. In the next chapter the successful generalization made by the noise-free training set is used to study the trend of the NNED method for the application to the evaluation of the segmentation errors when applied to similar noisy synthetic lesions.

References

Craven MW and Shavlik JW, "Using sampling and queries to extract rules from trained neural networks," in Machine Learning: Proceedings of the Eleventh International Conference, San Franscisco, CA, 1994.

Davis L, "Survey of edge detection techniques, computer Vision," *Graph. Image process.*, vol. 4, pp. 248-270, 1975.

Fan J and Yau DKY, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1454-1466, 2001.

Garcya-Silvente M, Garcya JA, et al., "A new edge detector integrating scale-spectrum information," *Image and Vision Computing*, vol. 15, pp. 913-923, 1997.

Hashem S, "Sensitivity analysis for feedforward artificial neural networks with differentiable activation functions," in Proceedings of the 1992 International Joint Conference on Neural Networks, IEEE Press, Piscataway, NJ, USA, 1992, vol. 1, pp. 419-424.

Ilardi P, "Neural Network Models of Sound Localization," BA thesis, Boston University, 1999.

Khashman A, "An Edge Detection Scheme Using Scale Space Analysis and Neural Network Arbitration," PhD thesis, University of Nottingham, Nott., UK, 1997.

Marr D and Hildreth E, "Theory of edge detection," *Proc. Royal Society of London B*, vol. 207, pp. 187-217, 1980.

Marr D and Poggio T, "A computational theory of human stero vision," *Proc. Royal Society of London B*, vol. 204, pp. 301-328, 1979.

Nakashima A, Hirabayashi A, et al., "Error correcting memorization learning for noisy training examples," *Neural Networks*, vol. 14, pp. 79-92, 2001.

Press WH, Flannery BP, et al., *Numerical Recipes in Pascal*: Cambridge University Press, 1989.

Prewitt J, "Object enhancement and extraction," in Picture Processing and Psychopictures, New York: Academic, 1970.

Rasband W. (2000). Scion image for windows. e-mail: wayne@codon.nih.gov. http://www.scioncorp.com

Roberts L, *Machine perception of three dimensional solids*: MIT Press, 1965.

Tarassenko L, *A guide to Neural Computing Applications*. London: NCAF, 1998.

van der Zwaag BJ, Slump C, et al., "Process identification through modular neural networks and rule extraction," in Computational Intelligent Systems for Applied Research: Proceedings of the 5th International FLINS Conference, (Ghent, Belgium, 16-18 Sept.), Sept. 2002, pp. 268-277.

van der Zwang BJ, Slump C, et al., "Process identification through modular neural networks and rule extraction," in Computational Intelligent Systems for Applied Research: Proceedings of the 5th International FLINS Conference, (Ghent, Belgium, 16-18 Sept.), Sept. 2002, pp. 268-277.

van der Zwang BJ and Slump K, "Analysis of neural networks for edge detection," in Proceedings of the ProRISC Workshop on Circuits, Systems and Signal processing, Veldhoven, the Netherlands, 2002, pp. 580-586.

van der Zwang BJ, Spaanenburg L, et al., "Analysis of neural networks in terms of domain functions," in Proceedings IEEE Benelux Signal Processing Symposium SPS-2002, (Leuven, Belgium, 21-22 March), March 2002, pp. 237-240.

Wong H, Caelli T, et al., "A model-based neural network for edge characterization," *Pattern Recognition*, vol. 33, pp. 427-444, 2000.

CHAPTER 5: COMPARISONS OF SKIN SEGMENTATIONS ERRORS

In this chapter we investigate the application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED). The aim is to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al [Xu, et al., 1999] is also used here as a comparison with other two proposed methods. These approaches are compared for synthetic lesions at different image signal to noise ratios (SNRs). The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be quantitatively and more accurately compared. Experiments are performed on digitized clinical photographs and also pigmented networks captured with the ELM technique. We demonstrate that we can enhance and delineate pigmented networks in skin lesions visually, and make them accessible for further analysis and classification (Section 5.3.2).

5.1 Introduction

The principle aim in the evaluation of a skin lesion image segmentation technique is to study the effectiveness in distinguishing a lesion from the background. For example, Hance [Hance, et al., 1996], Xu [Xu, et al., 1999] and Gao [Gao, et al.,

1998] have used expert clinical delineations as the gold standard to evaluate the automated segmented images. The disadvantage of using expert delineations is that it is dependent on human subjectivity. For example, Xu et al. [Xu, et al., 1999] in their study found that different clinical experts use different rules to segment an image. Some delineate the border where the lesion and non-affected skin meet. On the other hand, others locate the boundary where darker pigmentation and lighter pigmentation meet. An example of these variations made by four experts are contained in images provided by Xu et al. [Xu, et al., 1999]. When analysing the origin of these variations it was found that there are two main reasons for these, which could have a significant effect [Xu, et al., 1999]. The first was due to the clarity of lesion boundary i.e. the boundary is not always perfect and visually clear; it may have some width, as shown in Figure 5.1.

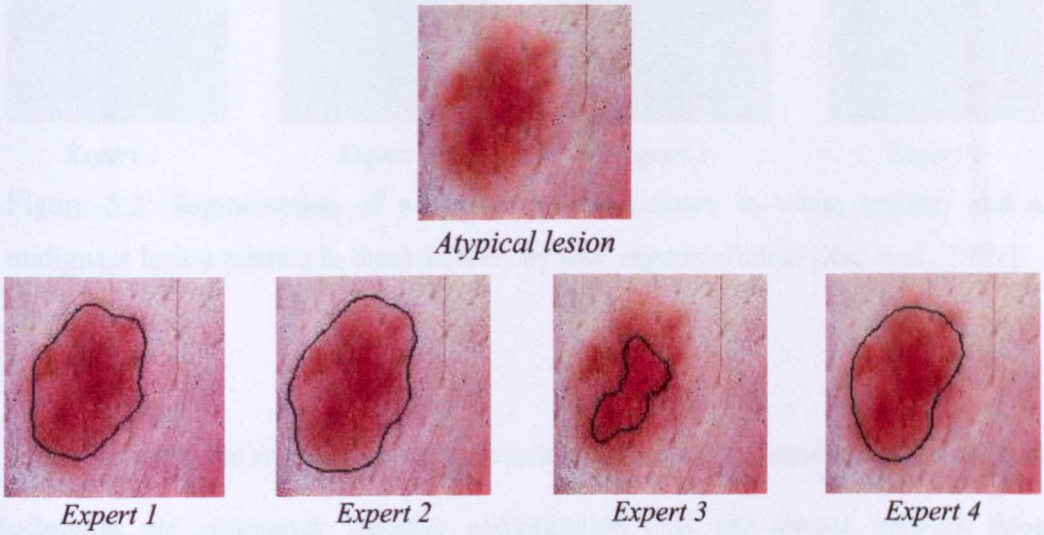


Figure 5.1: Segmentation of atypical lesion by four experts clinical [Xu, et al., 1999].

The second reason is the irregularity of the lesion boundary either benign or malignant: experts' segmentations for benign lesions have less variation than for malignant lesions, as shown in Figure 5.2 below:

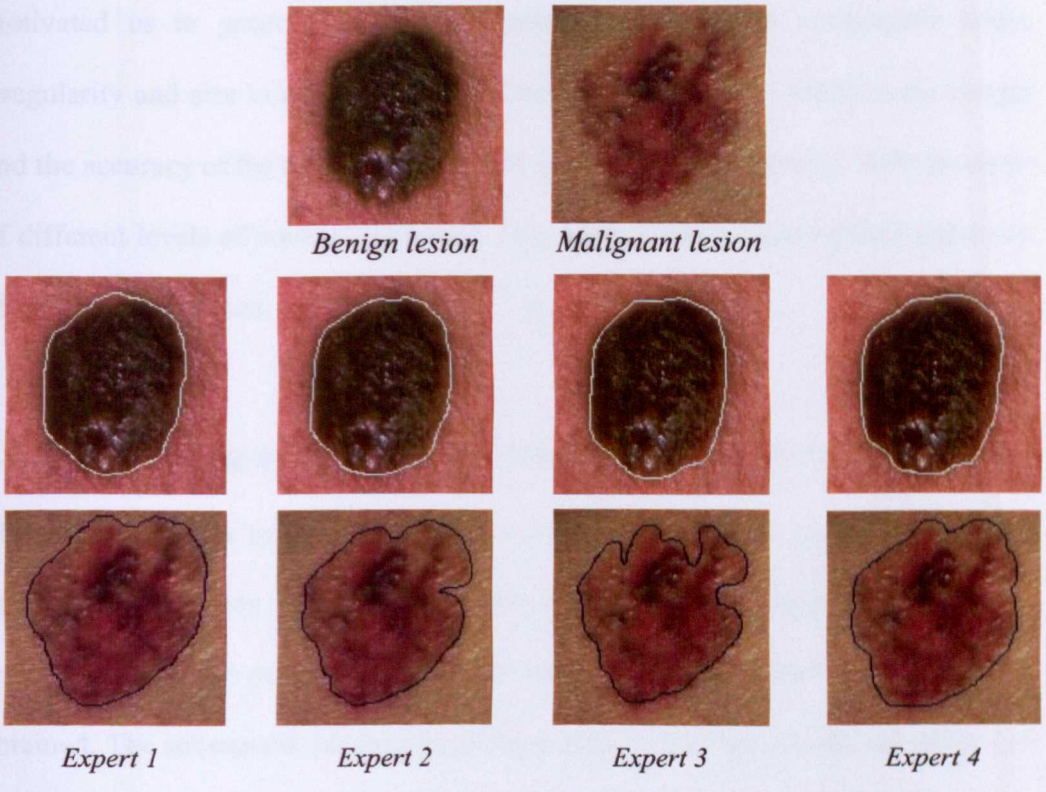


Figure 5.2: Segmentation of a benign lesions (shown in white border) and a malignant lesion (shown in black border) by four experts clinical [Xu, et al., 1999].

In this chapter, the two developed approaches (Chapter 3) to implement this type of technique are compared: iterative segmentation (IS) and neural network edge detection (NNED). These novel techniques are also compared with an established automatic skin segmentation method (ASS). The disadvantage of using expert

delineations of actual lesions to evaluate the algorithms is that this procedure is dependent on human subjectivity. In particular, the variation in the location of the lesion border between different experts is therefore not ideal when making a quantitative comparison between different automatic segmentation methods. This has motivated us to generate noise free synthetic images with comparable shape, irregularity and size to real skin lesions. Gaussian noise is then added to the images and the accuracy of the different techniques in segmenting the images in the presence of different levels of noise is evaluated. This enables a true quantitative comparison of different techniques.

5.2 Describing a skin lesion in terms of Gaussian noise

The design of noisy synthetic lesions (Chapter 3) was based on the assumption that colour transformation and image artefact removal can be easily achieved. After application of these processes, images of the form of the synthetic lesions will be obtained. The subsequent processing of these images will therefore be the subject of the comparisons in this chapter by three segmentation methods (IS, ASS, and NNED). Three sets of synthetic data are chosen to represent three different types of lesion irregularities. Since we know that segmentation methods may yield slightly different errors when applied to different lesion structures, then it is tedious and time consuming to generate numerous of similar structures of synthetic binary lesions as those shown in Figure 3.2. We have instead provided (a quite sufficient data) sets by corrupting three candidates synthetic images (shown in Chapter 3, Figure 3.2) with

different realisations of Gaussian noise (see Table 5.1). Figure 5.3 shows a sample of noisy synthetic lesions (for the case of very regular lesions),

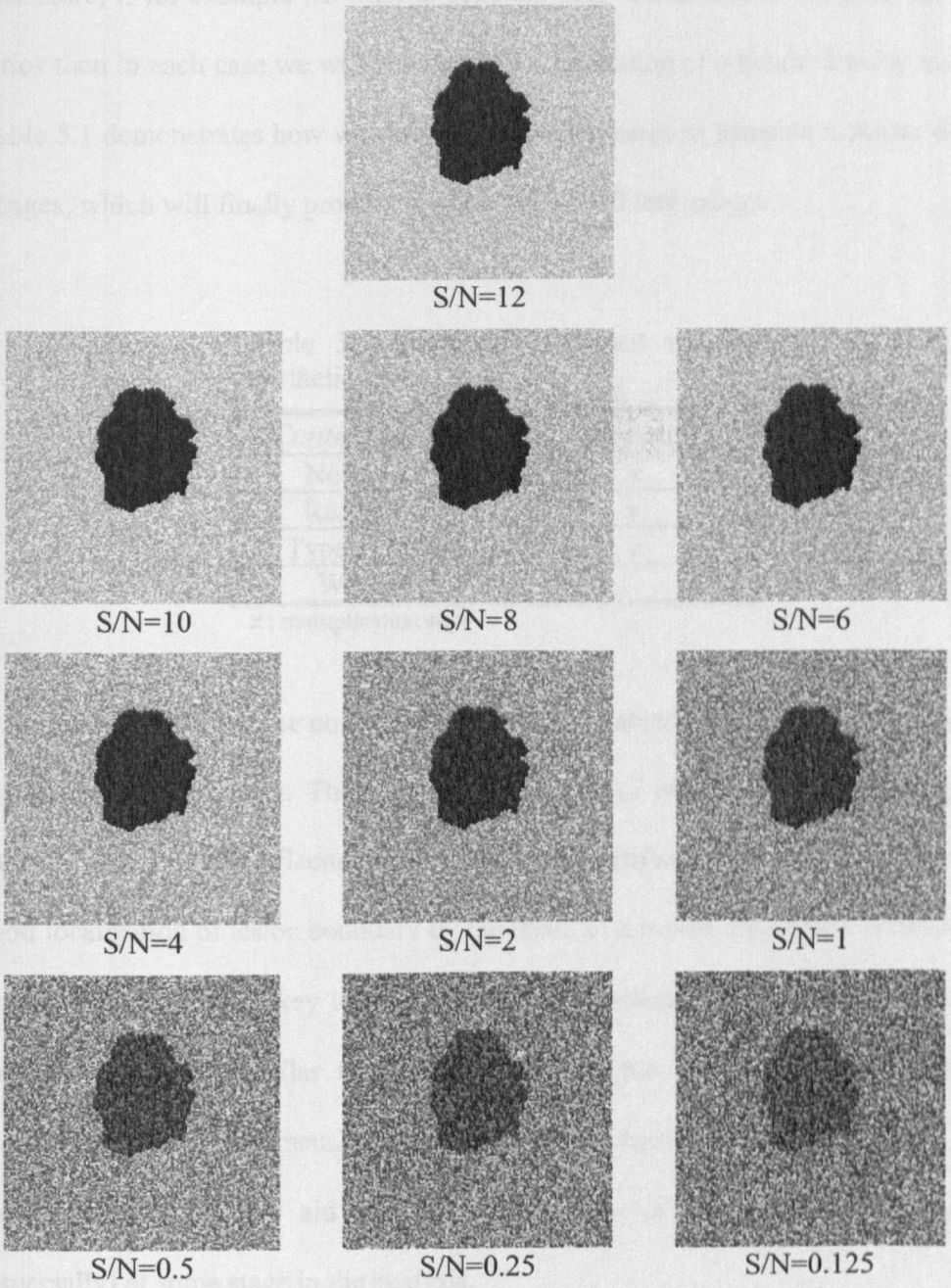


Figure 5.3: A very regular lesion corrupted with ten different realisations of Gaussian noise.

which are corrupted by different realisations of Gaussian noise. It should be noted that these images are only representing one out of ten different realisations. Therefore, if for example we want to generate each lesion case at ten different S/Ns ratios then in each case we will have a test set consisting of a hundred noisy lesions. Table 5.1 demonstrates how we can use these three cases to generate a cluster of test images, which will finally produce a whole set of 300 test images.

Table 5.1: Structure of a test set of synthetic lesions.

Contents of a set	*Quantity*
Noise levels	10 ×
Realizations	10 ×
Type of lesions	3 ×
Whole set	300

× : multiplication sign

It should be noted that the noise here may only characterize an approximation of the surrounding skin texture. There are a wide range of other noise elements in real lesions; such as light reflections, skin hair, noise artifacts...etc. The importance of good localization of lesion boundary by the result of a robust segmentation method is always considered as a key to the success of the evaluated method. Such a simple and general model similar to the one described here might provide a suitable preliminary test environment. Moreover, in the previous chapter it was shown how these synthetic images aid in the interpretation of different NNED outputs (especially) at some stage in the analysis.

In the following sections, results when segmenting both synthetic and real images are presented and discussed in more details.

5.3 Results

A quantitative comparison is provided by comparison of the techniques on the three references of synthetic lesions shown in Figure 3.2. However, we also provide a demonstration of the optimum technique on a real lesion.

5.3.1 Synthetic images

The segmentation methods are tested on the three synthetic lesions for a range of SNRs and are quantitatively evaluated using the error metric developed by Hance [Hance, et al., 1996]:

$$e = \frac{A \otimes B}{A} \tag{5.1}$$

where the exclusive-or output image represents the total area that does not overlap between the segmented lesion B and the original reference synthetic lesion A. A zero normalized error means that the evaluated segmented lesion B completely overlaps with the reference A, whilst an error of 1.0 is obtained when there is no overlap. This error measure is always between 0.0 and 1.0, and is independent of the size of a lesion. The error metric stated in Eqn. (5.1) has been used by Hance [Hance, et al., 1996], Xu [Xu, et al., 1999] and Gao [Gao, et al., 1998] to quantitatively evaluate the

result of segmentation. However in those cases the manual segmentation produced by a clinical expert is used as the gold standard.

Mean segmentation errors are computed and compared for the three synthetic lesions considered in this study (Figures 5.4 – 5.6), using twelve discrete levels of Gaussian noise corresponding to SNRs between 12 and 0.125 (which correspond to 10.8 dB and –9.0 dB respectively). It should be noted that the use of very low S/N in the simulation experiments to test the three segmentation methods is due to two main reasons. Firstly to test the three methods to their limits. Secondly it may be possible to investigate the application of these methods to segment other types of images in which the S/N is lower than the case of real skin lesion images. It is interesting to note that when trying to estimate the SNRs for the set of 18 real high quality SIAscope images (provided by Astron Clinica), it was found that the corresponding SNRs varied between 26.8 dB and 7.2 dB. However, lower SNRs are found in clinical photographs, for example in a set of 20 images provided by Xu et al.'s database of skin lesions; example images are shown in Figure 5.10. In Xu's database, the SNRs are found to vary between 13.1 dB and –6.7 dB.

Each individual point on the graphs uses ten realizations of a random noisy lesion (at the same SNR) to compute the mean segmentation error and its standard deviation. In the case of the IS method, two calculations are carried out, with and without the mapping function described in equation (3.6). In this way we can assess the importance of this procedure when applying the IS method.

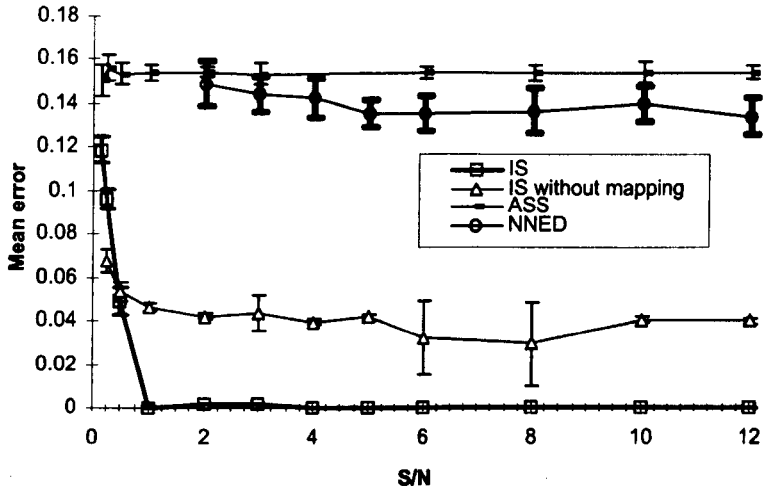


Figure 5.4: Comparison of mean error (%) between segmentation methods: *case of irregular lesion.*

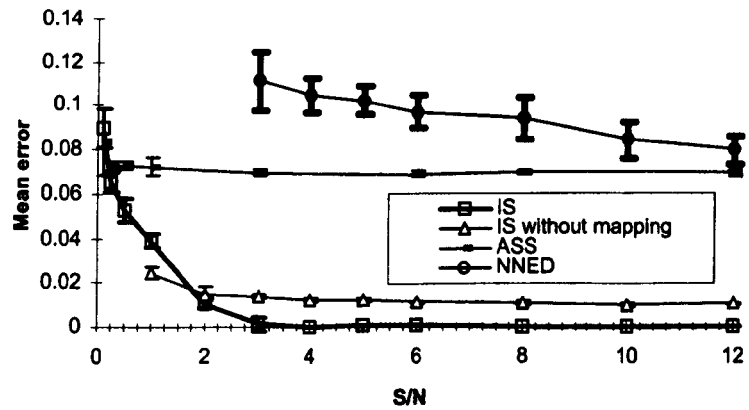


Figure 5.5. Comparison of mean error (%) between segmentation methods: *case of regular lesion.*

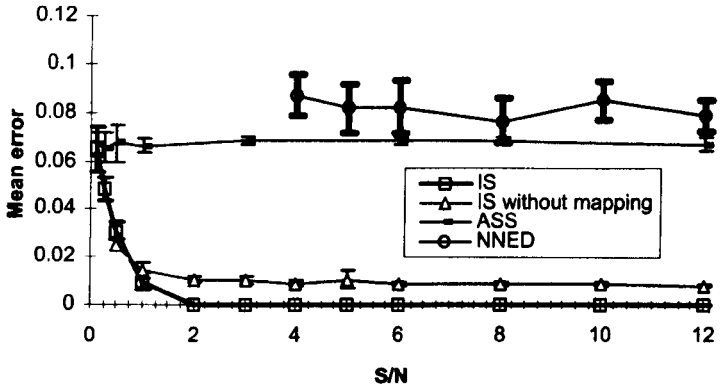


Figure 5.6. Comparison of mean error (%) between segmentation methods; *case of very regular lesion.*

At high signal to noise ratio (SNR) the IS algorithm retrieves the complete lesion segment with highest precision. On the other hand the ASS and the NNED methods have a significant level of error over the range of SNRs considered. At very low SNR, all techniques begin to break down but the IS still achieves the lowest segmentation error. *For synthetic lesions with different border irregularities the iterative segmentation method outperformed the others and should prove to be a valuable tool for skin cancer diagnosis.* In the next section the three methods will be tested on real skin lesions.

The IS method has been demonstrated to be the most effective method of segmenting the three types of synthetic lesions considered (Figures 5.4 to 5.6). One difference between the results from the ASS and IS methods is that the former approximates the

boundary region by fitting an elastic curve between refined edge pixels, from approximated lesion boundaries, whereas the latter does not. In the IS method, instead of approximating the lesion boundary, a subsequent object outlining operation is used to generate edge pixels from the resulting binary segment. The NNED result is also approximated by the same curve fitting as the ASS method. However, the NNED discrimination between edge and noise patterns is a critical issue especially for the case of very noisy lesions because the inputs to the NNED are the original noisy patterns, with no preprocessing like that used in the ASS and IS methods i.e. intensity mapping and noise filtering. This reduction of image operations along with the simple NNED structure used has an advantage in reducing the execution time of the NNED segmentation method.

5.3.2 Real lesions

Real lesion images may contain noise such as details from skin texture and hair that make it more difficult to localize the lesion border. In addition, the effect of undesirable color variations such as shadows and reflections (bright spots) tend to bias the color map e.g. when performing color segmentation [Hance, et al., 1996]. Therefore, when taking into account real images it is necessary to convert from a color image containing artifacts such as hair into a gray scale image with the artifacts removed. In the example demonstrated in Figure 5.8 we use the blue channel of the intensity of an RGB color skin lesion image. This approach has been demonstrated to provide the best results in global and dynamic thresholding algorithms [Ganster, et al., 2001]. Because real skin images often contain narrow objects such as hair and

other small objects, we have added a gray scale morphological opening [Gray, 1971] operation as the first step of data reduction. Opening is a succession of two operations, erosion and dilation. The erosion of gray intensities means that each pixel is replaced with the minimum value in the 3x3 neighborhood elements. Similarly, using the maximum value in a neighborhood performs grayscale dilation. The erosion has the effect of causing the thin and small objects to shrink or erode. Successively, dilation expands the eroded image by filling gaps in the lesion's contour and eliminating small holes. The remaining image operations are exactly the same as those explained in Chapter 3 (Section 3.3). Therefore, up to this stage, the general imaging tasks to delineate a colour lesion image could be summarized as shown in steps 1 to 5 in Table 5.1.

***Step1**: {Source image}*
Source image = Blue channel of {R,G,B} colour image
***Step2**: {Noise reduction}*
Grey morphology
Subtract median background noise
***Step3**: {Lesion enhancement}*
Map intensities with appropriate function
Smooth
***Step4**: {Optimal thresholding}*
Optimal thresholding
***Step5**: {delineate object}*
Outline binary object(s)
***Step6**:* {Object analysis}
Set minimum and maximum object size; *MinSize* and *MaxSize*
Scan the binary image *Until* *MinSize* < *Area* < *MaxSize*

Table 5.1. General algorithm steps to delineate colour lesion.

It has been noticed from experimental results when analysing various real images that, when subtracting background noise, the median operation (Step 2) has a major drawback of causing some holes to grow in an image. Therefore, to produce an image with a single lesion we have added another operation (Step 6). This process is useful when analysing an image to correct errors caused in the delineation process such as the delineation of images containing thick and dark hair. Scanning across the image is performed until a condition or a set of conditions is reached. For example, an image process, which is adopted here, is called "Particle Analysis" [Rasband, 2000]. This process will check a set of user entry options such as objects larger than a minimum size and less than a maximum size or to ignore any object touching edges of an image and also interior holes in an object. Figure 5.7 shows the entries of this process.

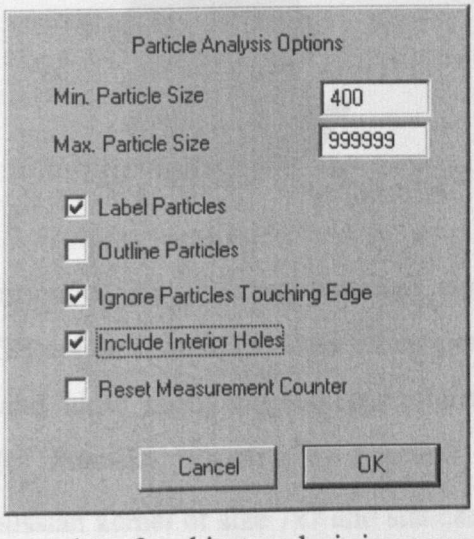


Figure 5.7: A user entry options for object analysis image process [Rasband, 2000].

Figure 5.8 shows the complete IS method applied to a real skin lesion. Fig 5.8(a) shows a selected colour malignant lesion image provided by Xu et al. database of

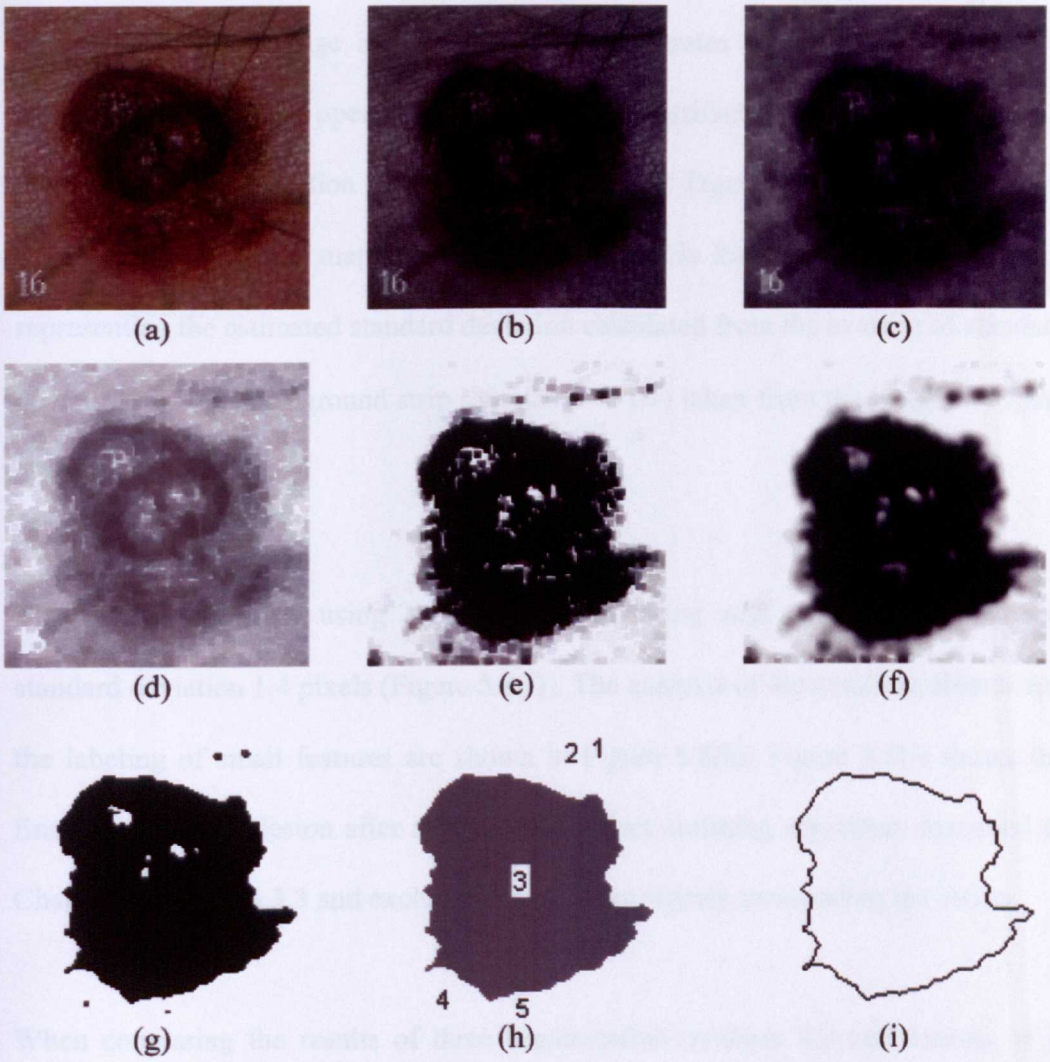


Figure 5.8. Demonstration of iterative segmentation algorithm. (a) Original pigmented lesion in RGB colour format. (b) Gray intensities of blue channel. (c) Application of open and close morphological operations. (d) Subtraction of median background. (e) Intensity mapping by function $F(\Phi)$ (Eqn. 3.4). (f) Smoothing by a 2D Gaussian kernel of size 7x7 and standard deviation 1.4 pixel. (g) Processing by isodata algorithm to produce a binary segment at an optimal threshold. (h) Analysis of the resulting objects. (i) Final segmentation result by excluding small objects less than *MinSize=400* pixels (excluding objects No. 1,2,4,5) and outlining image.

skin lesions [Xu, et al., 1999]. The image after selecting the blue channel is shown in Figure 5.8(b). The image in Figure 5.8(c) demonstrates the effect of applying the morphological opening operation to remove image artifacts and Figure 5.8(d) shows the image after subtraction of the background level. Figure 5.8(e) shows the image after application of the mapping function described in Eqn. (3.6) with σ =16.2803; representing the estimated standard deviation calculated from the average of standard deviations of two background strip samples ($w \times 10$) taken from the image in Figure 5.8(c).

This is then smoothed using 2D Gaussian smoothing with a 7x7 kernel size and standard deviation 1.4 pixels (Figure 5.8(f)). The analysis of the resulting objects and the labeling of small features are shown in Figure 5.8(h). Figure 5.8(i) shows the final border of the lesion after applying the object outlining algorithm described in Chapter 3, section 3.3.3 and excluding small noise objects surrounding the lesion.

When comparing the results of three segmentation methods for real lesions, it is found that there is consistency in the relative performances of these methods when applied to synthetic lesions. To illustrate this, Figure 5.9 shows how the IS, ASS, and NNED segmentation results are similar for both real and synthetic lesions. The similarity between the two categories of results indicates the success of the synthetic model. For example, the IS method could preserve better the lesion's topology in both real (Figure 5.9(d)) and synthetic (Figure 5.9(g)) lesions than the other two methods. Approximating the experts' delineations was one of the aims of the method

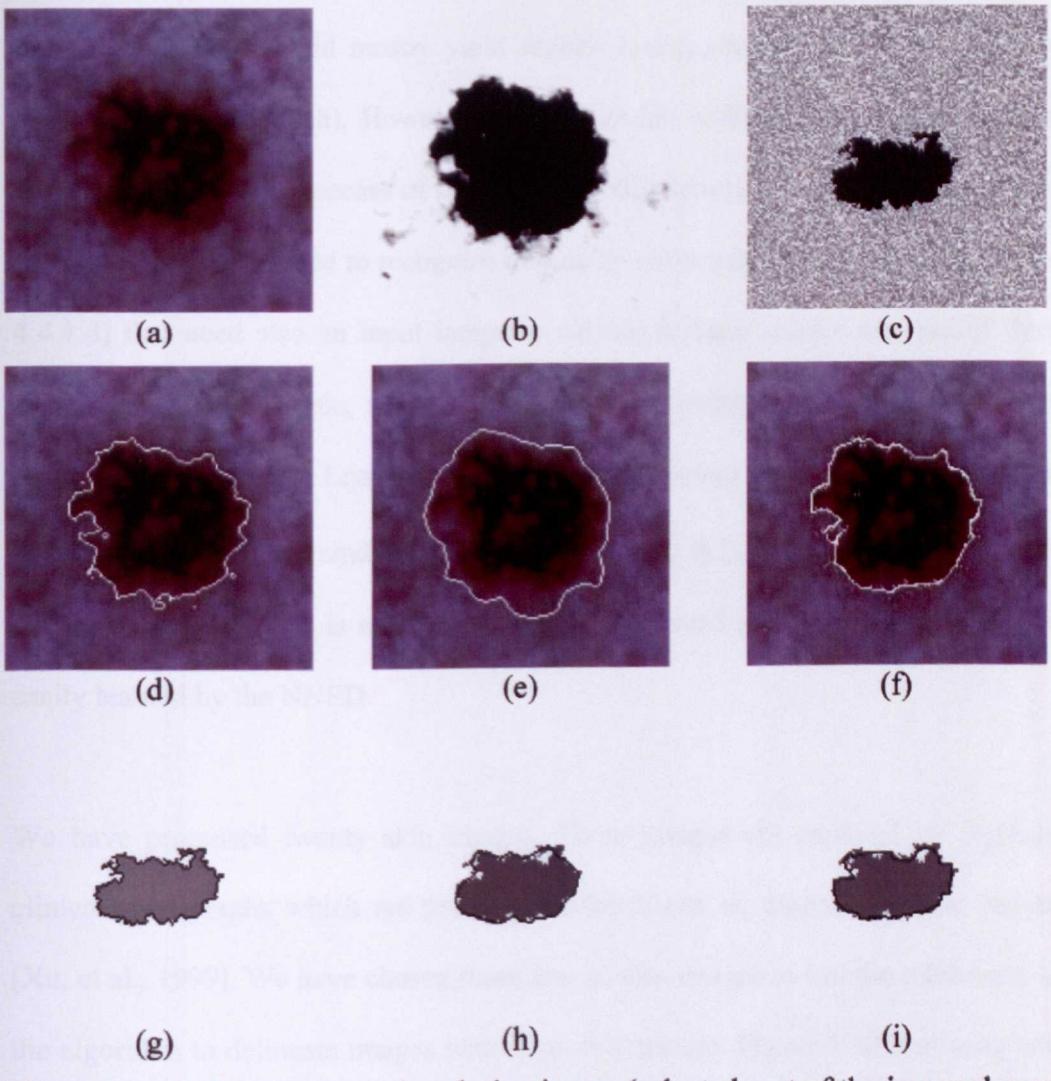
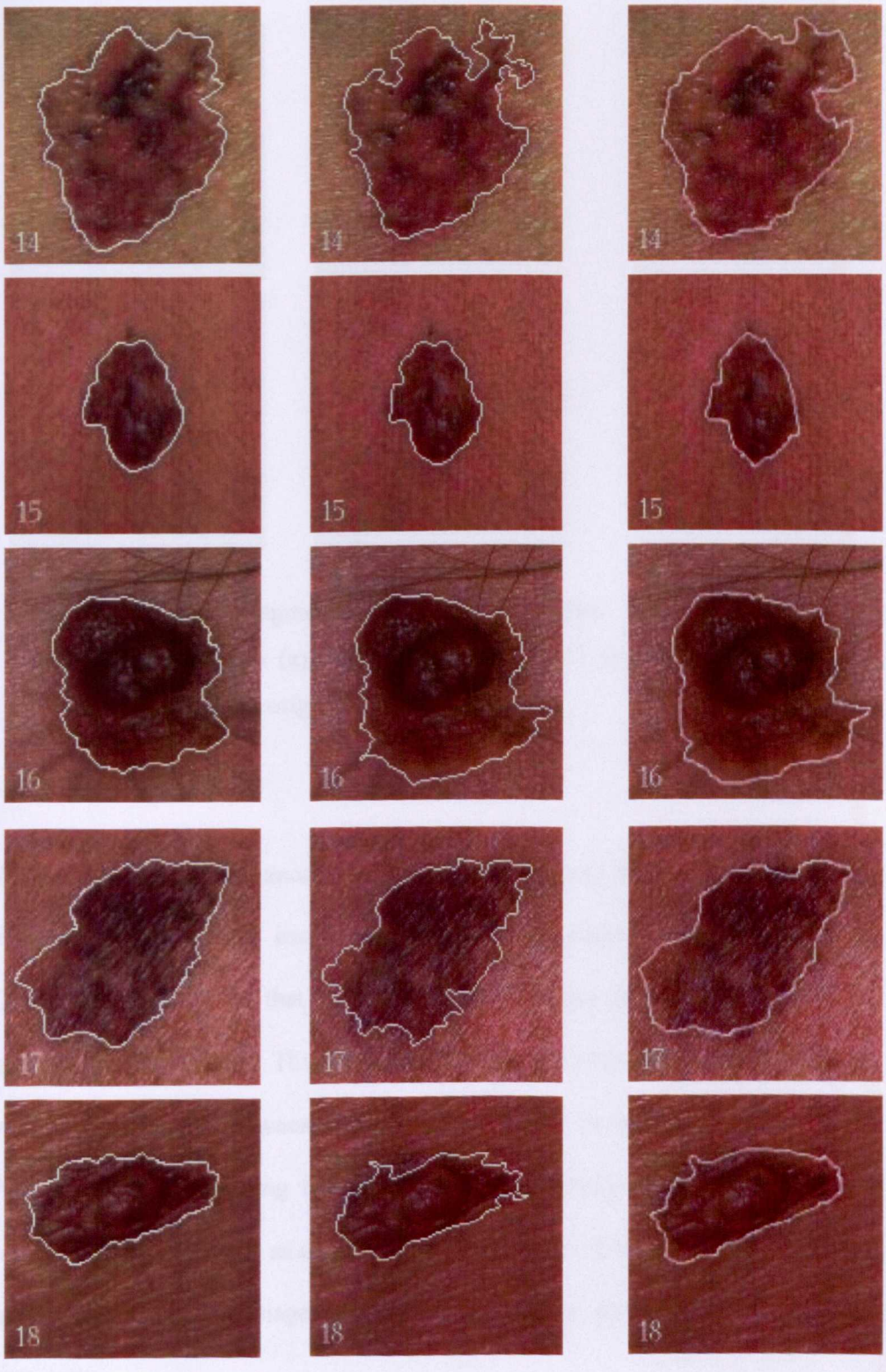
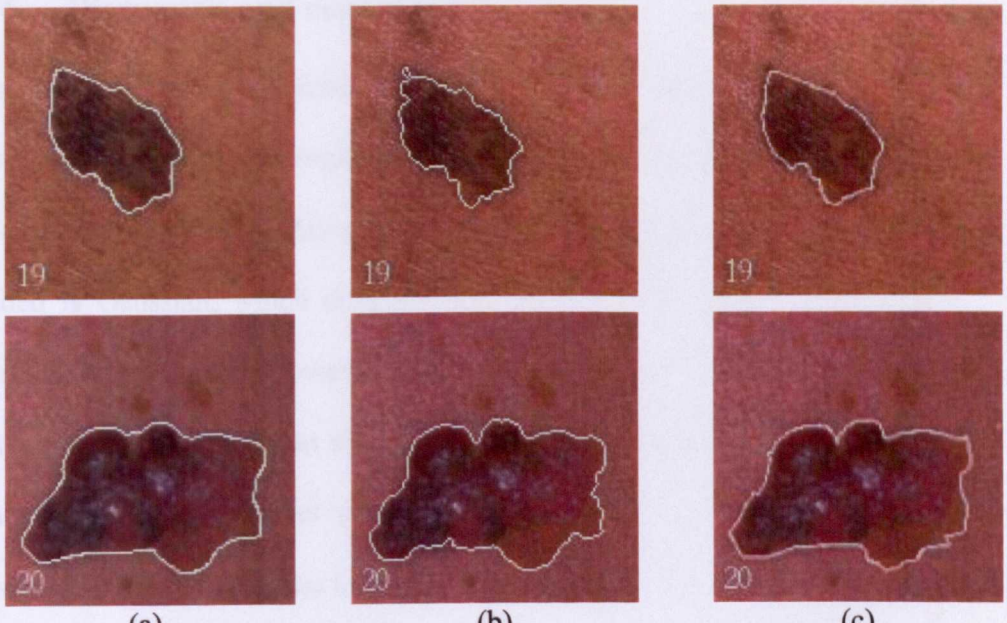


Figure 5.9. (a) A SIAscope colour lesion image (selected part of the image shown in Figure 3.1(b)). (b) Intensity mapping by function $F(\Phi)$ (Eqn. 3.4). (c) Irregular noisy synthetic lesion (S/N=3.0). (d)-(e) are the lesion boundaries produced by IS and ASS, respectively for the colour lesion in (a). (f) NNED segmentation result for intensity mapped image in (b). Similarly, (g) to (h) are the output results for IS, ASS, and NNED, respectively, when applied to synthetic lesion in (c).

of Xu et al. and accurate localization of a lesion boundary was of less importance in this method. This would mostly yield regular lesion structures as those shown in Figures 5.9(e) and 5.9(h). However, the NNED has a different method to localize edge boundaries. The success of the trained NNED here is an image dependent, i.e. the NNED is only trained to recognize non-noisy sharp edges (see Chapter 4, section 4.4.1.4) that need also an input image to contain at least similar patterns of these sharp edges. For example, when a simple intensity mapped image in Figure 5.9(b) (using $F(\Phi)$ function in Eqn. 3.4) of the real lesion shown in Figure 5.9(a) is applied to NNED, the output boundary produced in Figure 5.9(f) is similar to one obtained by the IS method. This is mainly because the enhanced intensity mapped image is easily learned by the NNED.

We have processed twenty skin images. These images are captured by digitised clinical photographs which are provided by the Xu et al. database of skin lesions [Xu, et al., 1999]. We have chosen these low quality images to test the robustness of the algorithm to delineate images with clear skin texture. Figure 5.10 also compares more outputs results from the three segmentation methods when they are applied to malignant lesions; produced by Xu et al.'s skin lesion database [Xu, et al., 1999]. It is clearly shown that the IS method is better at preserving most of the lesion's topology than the other two methods.


14
14
14
15
15
15
16
16
16
17
17
17
18
18
18



(a) (b) (c)

Figure 5.10: Final segmentation results (boundary delineations) for seven malignant images with (a) ASS, (b) IS, (c) NNED methods are superimposed with the corresponding original colour images.

We have also processed another set of eighteen images, which are captured with the ELM technique. These images are provided by Astron Clinica (section 3.1.2, Chapter3). It was found that the ASS method produces relatively poor results when applied to these images. This is mainly because the class of ELM images contains very low variations between intensities, which will be incorrectly mapped by the ASS method. Its mapping function will include details of the background in the resulting mapped image, as discussed in section 3.3.1, Chapter 3. However, the high quality of this class of images makes the segmentation by IS and NNED methods an easy task

5.4 Discussion and conclusion

The synthetic lesions utilized have enabled a truly quantitative comparison of the three techniques as the segmented images can be compared with the originals. Clearly the drawback of this approach is that the images are not colour and the noise added is Gaussian, which does not account for image artifacts such as hairs or surface bright spots. However colour transformation and artifact removal are established methods within this field. We have therefore assumed that after color transformation and artifact removal of real images we will obtain lesions that resemble the noisy synthetic lesions e.g. shown in Figure 5.3. We have evaluated the three segmentation techniques for the subsequent processing of such images.

The initial segmentation in the IS algorithm uses an accurate and comprehensive thresholding technique that iteratively analyses the whole image, object and background, provided that both the lesion and its background have distinct average gray levels. In contrast the thresholding technique adopted in the ASS method is mostly suitable for the case of a lesion having distinct gray levels from that of background. In practice this case may not always exist especially when considering a wide range of lesion scenes with widely different properties [Ridler and Calvard, 1978]; e.g. effect of noise and image details such as skin texture and hair. Moreover, the subsequent initial thresholding only finds an approximate lesion boundary, which does not always represent true region boundaries even if this region is refined using edge information in the image [Xu, et al., 1999].

Even though the curve fitting represents a slight error, the thresholding technique used in the ASS method is the main source of error. This can be proved, as Figure 5.11 shows, by curve fitting to a noise free, very irregular lesion. The error between the actual lesion area of Figure 5.11(a) and the lesion after curve fitting (close up section shown in Figure 5.11(c)) is approximately 1%.

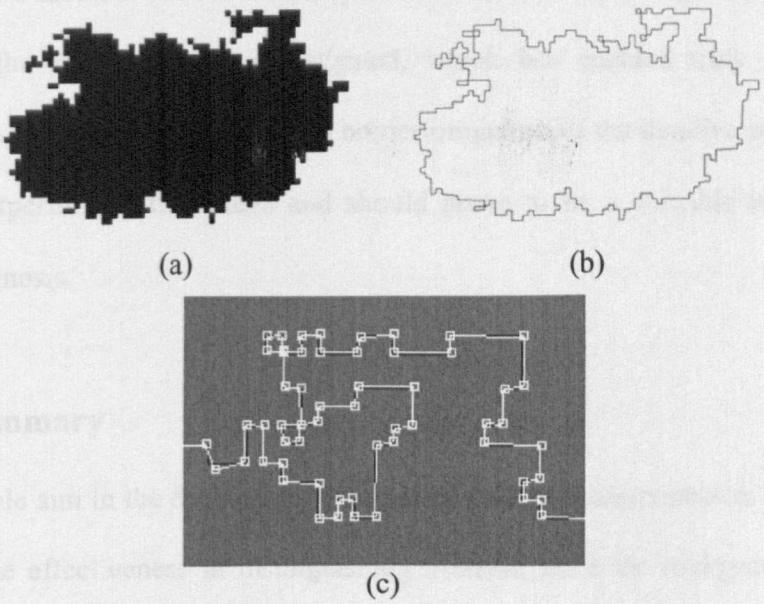


Figure 5.11. (a) Synthetic lesion of Figure1(a) (magnified X10). (b) Object outline of lesion (a). (c) a close up section clip of the result of curve fitting (top right corner of the lesion) showing a new curve fitting (white) closely fitting the original lesion.

We have established the potential of the iterative segmentation method when segmenting synthetic lesions with added Gaussian noise and demonstrated how this can be applied to real images. In future we intend to proceed in two directions: firstly, the visual enhancement and delineation of skin lesions can make them

accessible to further analysis and classification (e.g. using the IS method) and, secondly, to develop more sophisticated noise models to incorporate image artifacts, which will also aid in the comparison of the various skin segmentation methods.

In conclusion, we have developed two new methods (iterative segmentation and neural network edge detection) for segmentation of skin lesions, and compared them with the ASS method. We have used synthetic lesions in analyzing the segmentation errors of the three methods investigated, which has enabled truly quantitative comparison. For lesions with different border irregularities the iterative segmentation method outperformed the others and should prove to be a valuable tool for skin cancer diagnosis.

5.5 Summary

The principle aim in the evaluation of a skin lesion image segmentation technique is to study the effectiveness in distinguishing a lesion from the background. In this thesis, two new approaches to implementing this type of technique are suggested: iterative segmentation (IS) and neural network edge detection (NNED). These novel techniques are also compared with an established automatic skin segmentation method (ASS). The aim is to quantitatively analyze the error in locating the border due to the application of an automated segmentation method. The disadvantage of using expert delineations of actual lesions to evaluate the algorithms is that this procedure is dependent on human subjectivity. In particular, the variation in the location of the lesion border between different experts is therefore not ideal when

making a quantitative comparison between different automatic segmentation methods. This has motivated us to generate noise free synthetic images with comparable shape, irregularity and size to real skin lesions. Gaussian noise is then added to the images and the accuracy of the different techniques in segmenting the images in the presence of different levels of noise is evaluated. This enables a true quantitative comparison of different techniques.

Mean segmentation errors are computed and compared for the three synthetic lesions considered in this study. At high signal to noise ratio (SNR) the IS algorithm retrieves the complete lesion segment with highest precision. On the other hand the ASS and the NNED methods have a significant level of error over the range of SNRs considered. At very low SNR, all techniques begin to break down but the IS still achieves the lowest segmentation error. The IS technique is also tested on an image of a real skin lesion. *In conclusion, for synthetic lesions with different border irregularities the iterative segmentation method outperformed the others and should prove to be a valuable tool for skin cancer diagnosis.*

References

Ganster H, Pinz A, et al., "Automated Melanoma Recognition," *IEEE Trans. on Medical Imaging*, vol. 20, no. 3, pp. 233-239, 2001.

Gao J, Zhang J, et al., "Segmentation of dermatoscopic images by stabilized inverse diffusion equations," in Proceedings of the International Conference on Image Processing '98, IEEE, Piscataway, NJ, 1998, vol. 3, pp. 823-827.

Gray SB, "Local properties of binary images in two dimensions," *IEEE Trans. Computer*, vol. 20, no. 5, pp. 551-561, 1971.

Hance GA, Umbaugh SE, et al., "Unsupervised color image segmentation with application to skin tumor borders," *IEEE Eng. Med. Biol. Mag.*, vol. 15, no. 1, pp. 104-111, 1996.

Rasband W. (2000). Scion image for windows. e-mail: wayne@codon.nih.gov. http://www.scioncorp.com

Ridler TW and Calvard S, "Picture thresholding using an iterative selection method," *IEEE Trans. on Systems, Man and Cybernetics*, 1978.

Xu L, Jackowski M, et al., "Segmentation of skin cancer images," *Image Visions Computing*, vol. 17, pp. 65-74, 1999. http://www.cs.wright.edu/people/faculty/agoshtas/paper_fig.html

CHAPTER 6: CONCLUSIONS

The work presented in this thesis has introduced two novel approaches to edge detection. The techniques have been applied to skin lesion segmentation, and a comparison has been made with an existing technique. The first method was a mainly thresholding segmentation method where an optimal threshold is determined iteratively by an isodata algorithm. The second method proposed was based on neural network edge detection and a rational Gaussian curve that fits an approximate closed elastic curve between the recognized neural network edge patterns. A quantitative comparison of the techniques was enabled by the use of simulated lesions to which Gaussian noise was added. The proposed techniques were also compared with an established automatic skin segmentation method. It was demonstrated that for lesions with a range of different border irregularity properties the iterative segmentation (IS) method provides the best performance over a range of signal to noise ratios. The IS method was also demonstrated to have similar performance when tested on real skin lesions (Chapter 5). In the subsequent discussion, each chapter will be concluded individually and then overall conclusions and suggestions for further work will be made.

Early detection is the most effective method to reduce mortality from malignant melanomas because generally the survival rate is inversely proportional to the thickness of the lesion [Lee, et al., 2003]. In an attempt to combat the rapidly increasing incidence rate of melanomas, many physicians advocate some kind of

automatic early diagnostic aided systems to improve diagnosis [Lee, et al., 2003]. Important clinical features contribute in the success of many classifiers [Binder, et al., 2000;Colot, et al., 1998;Ercal, et al., 1994;Guthowicz-Krusin, et al., 1997;Hall, et al., 1995;Schindewolf, et al., 1993;White[a], et al., 1991;White[b], et al., 1991]. In this thesis, the research has been focussed on border irregularity which is an important clinical feature that is used when assessing the malignancy of the skin lesion.

Chapter one described the motivation of the thesis, which is the accurate automatic delineation of melanoma lesion border shape, which would improve the effectiveness of a computer-aided system for the analysis and classification of benign or malignant lesions. The vast field of image processing, along with the issues concerning image segmentation, were highlighted. The particular problem of the segmentation of melanoma images was described. This chapter also presented the objectives of the work to be carried out and an overview of the structure of the thesis.

In chapter two, the theoretical background and a literature review were presented. The application of Neural Networks (NN) was described. NNs are trying to mimic or simulate the way a simple biological nervous system is believed to operate. The models used by NNs are capable of solving sophisticated, perhaps "intelligent", computation similar to those that the human brain routinely performs. NNs were investigated as one of the key elements used in this research. There are numerous kinds of NNs and also a wide range of their applications. A discussion of the applications of NNs to edge detection was presented. The most widely used NNs

architecture, the back propagation (BP) training algorithm, was chosen in this work and the learning algorithm was described. Finally, the mathematical background of another edge detection technique, optimal thresholding, was described which was the basic building block of the developed IS algorithm.

In chapter three, the three techniques that are applied in this research to the segmentation of pigmented skin lesions were described. We investigated the application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED). The aim was to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al. [Xu, et al., 1999] was also used here to verify the other two proposed methods. These three approaches were compared for synthetic lesions at different image signal to noise ratios (SNRs) as discussed in Chapter 5.

In chapter four, a new technique for generating a neural network training set was considered for edge detection. A limited number of noisy edge patterns, was proposed to analyse the capability of neural networks edge detection for both synthetic images and real images. Various methods were applied to reduce the neural network noisy training set; e.g. variation of training set size. However, the use of the *S/N* ratio, rather than the standard deviation of the noise, in characterising the noisy patterns in the training set was one of the successful methods which have been considered in this study to modify the technique of generating additive noise to the

training set. When constructing and experimenting on NNs with various noise free training sets, the total number of prototype edge patterns and their redundancy are considered. A remarkable improvement was shown in the neural network recognition when the total number of noise free prototype edge patterns is forty-two incorporating five edge profiles with the possible orientations of each edge profile.

We have also analysed the weights of a neural network edge detector's hidden nodes. This has given an easy insight into the internal functionality (e.g. low-pass or averaging behaviour, Chapter 4) of the neural network as an edge detector [van der Zwaag, et al., Sept. 2002]. The weights between hidden units and the output unit represents the importance of the hidden unit's edge detection outcome. Some larger NNs have also been trained and analysed. The larger the network, the more variety in behaviour among the neural units. In a few cases, certain units showed very strong high-order behaviour, indicating that those units functioned as noise detectors only. Further work, using this analysis method, can be reached by analysing the behavior of different neural network architectures; e.g. RBF (Radial Basis Function) Networks.

In chapter five, we investigated the application of two approaches to the skin lesion segmentation problem; iterative segmentation (IS) and neural network edge detection (NNED). The aim was to quantitatively analyse the error in locating the border due to the application of an automated segmentation method. The automatic skin segmentation (ASS) method presented by Xu et al [Xu, et al., 1999] was also used

here as a comparison with other two proposed methods. These approaches were compared for synthetic lesions at different image signal to noise ratios (SNRs). The use of synthetic lesions is advantageous in initial analysis and verification, as by knowing the true position of the lesion border the different methods can be quantitatively and more accurately compared. Expert delineation of actual lesions was not chosen to evaluate these algorithms, because this procedure is dependent on human subjectivity (e.g. Figure 5.1). Experiments were performed on digitized clinical photographs and also pigmented networks captured with the ELM technique. We demonstrated that we could enhance and delineate pigmented skin lesions visually, and make them accessible for further analysis and classification (Section 5.3.2).

The conclusions of the individual chapters have been presented. The significance of the work in this thesis has been shown. It is interesting to sum up that the proposed objectives (listed in Chapter 1) have been successfully achieved. The first objective was to provide the appropriate preliminary experiments for the investigated methods (ASS, IS, and NNED). The preliminary experiments on synthetic images have provided an excellent environment to verify the edge detection capability of these methods, e.g. has a major effect in designing the algorthim for the IS method. The second objective, which was to design and verify the NNED method, was met through the application of various experiments to analyse the problem of neural network edge detection (NNED) for both synthetic and real life images. Moreover, the success achieved in the generalised NNED was used as the basis in the design of

the NNED method (Chapters 4 and 5). The third objective, which was to analyse quantitatively the error in locating the border due to the application of the three segmentation methods (ASS, IS, and NNED), was achieved through the implementation of the three methods on the noisy synthetic images. In addition, the quantitative and accurate comparisons of the results obtained from ASS, IS, and NNED segmentation methods, have indicated the accuracy of these methods. The final objective was to verify the most appropriate method, for the class of real skin lesion images, which could characterize the structure of lesion images and then aid for accurate measurements and classification. This was met through the application of the three investigated methods to both clinical photographs and ELM real images.

For lesions with different border irregularities the iterative segmentation (IS) method outperformed the others and should prove to be a valuable tool for skin cancer diagnosis. However, there is a need to expand the domain of application of the IS method to numerous types of malignant lesions so as to extract different features and make them available for a classifier. Further work in this area could proceed in three directions. Firstly, the visual enhancement and delineation of skin lesions can make them accessible to further analysis and classification (for example using the IS method), secondly, to develop more sophisticated noise models to incorporate image artifacts, which will also aid in the comparison of the various skin segmentation methods and, lastly, to investigate the application of these developed methods to different classes of clinical images. One example, is the analysis and segmentation of mammograms [Petroudi and Brady, 2003].

References

Binder M, Kittler H, et al., "Computer aided epiluminescence microscopy of pigmented skin lesions: the value of clinical data for the classification process," *Melanoma Research*, vol. 10, no. 6, pp. 556-561, 2000 .

Colot O, Devinoy R, et al., "A color image processing method for melanoma detection," in The first International medical Imaging Computing and Computer-Assisted Intervention, Cambridge, MA, 1998, pp. 562-569.

Ercal F, Chawla A, et al., "Neural network diagnosis of malignant melanoma from color images," *IEEE Trans. on Biomed. Eng.*, vol. 41, no. 9, pp. 837-845, 1994 .

Guthowicz-Krusin D, Elbaum M, et al., "Can early malignant melanoma be differentiated from atypical melanocytic nevus by in vivo techniques? Part II. Automatic machine vision classification," *Skin Research and Technology*, vol. 3, pp. 15-22, 1997 .

Hall PN, Claridge E, et al., "Computer screening for early detection of melanoma- is there a future?," *British Journal of Dermatology*, vol. 132, pp. 325-338, 1995 .

Lee TK, McLean DI, et al" ‹.Irregularity index: A new border irregularity measure for cutaneous melanocytic lesions," *Medical image analysis*, vol. 7, no. 1, pp. 47-64, 2003 .

Petroudi S and Brady M, "An automated algorithm for breast background segmentation," in Medical Image Understanding and Analysis 2003, University of Sheffield, UK, 2003, pp. 181-184.

Schindewolf T, Stolz W, et al., "Classification of melanocytic lesions with color and texture analysis using digital image processing," *Analytical and Quantitative Cytology and Histology*, vol. 15, no. 1, pp. 1-11, 1993 .

van der Zwaag BJ, Slump C, et al., "Process identification through modular neural networks and rule extraction," in Computational Intelligent Systems for Applied Research: Proceedings of the 5th International FLINS Conference, (Ghent, Belgium, 16-18 Sept.), Sept. 2002, pp. 268-277.

White[a] R, Rigel DS, et al., "Computer applications in the diagnosis and prognosis of malignant melanoma," *Dermatologic Clinics*, vol. 9, no. 4, pp. 695-702, 1991 .

White[b] RG, Perednia DA, et al., "Automated feature detection in digital images of skin," *Computer Methods and Programs in Biomedicine*, vol. 34, pp. 41-60, 1991 .

Xu L, Jackowski M, et al., "Segmentation of skin cancer images," *Image Visions Computing*, vol. 17, pp. 65-74, 19 .99
http://www.cs.wright.edu/people/faculty/agoshtas/paper_fig.html

APPENDIX A – PUBLICATIONS

Rajab MI and Woolfson MS, "Segmentation of dermatoscopic images by iterative segmentation algorithm," in Proceedings of Medical Image Understanding and Analysis 2003, The University of, Sheffield, 2003[a], pp. 121-124.

Rajab MI and Woolfson MS, "Analysis of neural network edge detection," in WSEAS Transactions on Systems; 4[th] WSEAS International Conference on Neural Networks and Applications (WSEAS NNA 2003), WSEAS, Vouliagumeni,Athens, Greece, 2003[b], vol. 2 Issue(3), pp. 649-654.

Rajab MI, Woolfson MS, et al., "Application of region based segmentation and neural network edge detection to skin lesions," *Computerized Medical Imaging and Graphics*, to be published.