DOCTOR OF PHILOSOPHY

Texture and colour for automatic image-based skin lesion analysis.

Round, Andrew John

Award date:
1998

Awarding institution:
Bangor University

Texture and Colour for Automatic Image-Based Skin Lesion Analysis

Andrew John Round

Thesis submitted in candidature for the degree of

Doctor of Philosophy

August 1998

School of Electronic Engineering and Computer Systems
University of Wales, Bangor
United Kingdom

Acknowledgements

My sincere thanks and gratitude go to Dr Andrew Duller and Peter Fish for their invaluable advice and guidance throughout the course of this research. I would also like to acknowledge the other staff and postgraduate students at Bangor who have given support and encouragement.

Special thanks also go to my family, to Tazmin Campbell, and to many friends for their patience, understanding and motivation.

Summary

The research presented here considers automatic diagnosis support for skin cancer. The role of computer-based diagnosis, and its value within a primary care situation are examined resulting in synthesis of aims, requirements and properties for an effective system — a system based on digital optical images captured and processed using low-cost commercial computer technology.

The issues involved in acquisition of lesion boundaries are discussed. The value of accurate and robust boundaries, in terms of both directly obtainable diagnostic features and in enabling lesion property evaluation, is identified. Previous research has proposed the edge focusing process. This work has addressed the improvement, in terms of potential for future development, evaluation and reuse, of this process through porting it to a highly modular form in the Khoros environment.

The role of colour analysis and its value in terms of provision of diagnostically useful features is investigated, and the central importance of segmentation is identified. The fundamental properties of effective segmentation of lesion image colours are identified as a need to reflect human perception of colour similarity and a basis on local regions. A new region-based segmentation technique using data transformed to a perception-uniform colour-space is presented and shown to yield promising results.

Finally the use of texture information is discussed. The nature and properties of the large-scale texture of skin patterning and its disruption are investigated and an abstracted representation proposed. A new technique is presented and shown to be effective in extracting the qualities of the skin patterning. Methods for analysing this representation of the patterning to quantify the disruption attributable to the lesion are proposed and developed. The combination of these extraction, analysis and disruption evaluation techniques is shown to be effective in relation to both visual assessment of disruption and diagnostic performance.

Contents

Chapter 1

Introduction

1.1 Motivation

The incidence of skin cancer is rapidly increasing throughout the world and currently rivals that of all other forms of cancer put together [1–3]. Malignant melanoma is the most lethal of the skin cancers, accounting for only about 5% of cases but 80-85% of fatalities [2–4].

Early recognition and treatment of skin cancer is vital as malignant lesions, especially invasive melanomas have a much increased fatality rate as the lesions develop [1, 5–7]. Simple surgical excision is well known to be highly effective where lesions are treated early [1, 3–5] and can effect a complete cure [8]. Effective diagnosis must therefore be widely available, rapid and able to discriminate using early indicators.

Histological analysis is often regarded as the only truly reliable method for establishing the nature of a lesion [6], however the use of this invasive technique for all lesions is impractical in terms of time, cost and patient inconvenience. Identification and differential diagnosis of skin lesions is normally carried out visually, the skin being naturally readily accessible to this simple form of inspection. Accuracy of such diagnosis is however not assured, difficulties are caused as each lesion type does not present a unique appearance [5]. Experienced dermatologists are

generally not more than 65-70% accurate for non-typical pigmented lesions [5, 9] and the likely first point of contact for a patient, the GP, is unlikely to match this accuracy [9, 10], the broad expertise required of a GP naturally excluding the specific depth required for skin lesions.

Early recognition has been pursued through campaigns aimed at GPs and the general public increasing awareness of signs and aiming to reduce the time taken before presentation to a GP and referral to a dermatologist. The simplicity and brevity required for the information content of such campaigns however entails a limit on the possible accuracy, as does the importance of avoiding false indications of cancers as benign. In addition, the subjective quality inherent in human assessment of indicators naturally results in inter and intra observer inconsistencies and consequently a further reduction in accuracy.

There is a need then, for a means of effectively servicing first presentations of 'suspect lesions' both quickly and accurately and with minimal inconvenience to the patient. A computer based system could provide a means of encapsulating the specifics of lesion diagnosis, and automated acquisition of diagnostic indicators will provide consistency. The majority of quantifiable, and non-invasively available, indicators currently used in lesion diagnosis are available in an image of the lesion. A computer system based on standard digital video capture technology has the advantages of both low capital and running costs, as well as causing a minimum of patient inconvenience.

1.2 Previous Research

The value of an automatic diagnosis support system for skin lesions is reflected in the considerable research interest in this area — particularly concerning lesion boundary identification and colour analysis.

Many of the important diagnostic indicators identified for the clinical assessment of malignancy in melanoma can be obtained from boundary information alone. Well known indicators in this category include size, border irregularity, notching and asymmetry. The extent of the lesion is also a vital prerequisite for analysis of

features of the lesion area and for comparative measures between properties of the lesion and the surrounding skin. Much of the current research has consequently concentrated on techniques for boundary identification and a wide variety of techniques have been proposed (e.g. [11–15]). As part of a previous study here at Bangor, a process has been developed [16, 17] which can provide an accurate boundary in a manner robust to many image imperfections.

The importance of colour in diagnosis of skin lesions is widely recognized. Not only does colour provide valuable information for the identification of lesion boundaries, colour features, such as the presence of certain shades and variability of lesion colour, also provide vital diagnostic information, as reflected in their inclusion in both skin cancer checklists and in differential diagnosis descriptions. Colour has consequently been important in research into automated diagnosis both as part of boundary identification methods and in detection of colour features.

Texture analysis has not attracted the same level of interest as have boundaries and colour, perhaps because such features do not commonly appear in skin cancer public information. Texture information can be considered at different scales; small-scale texture where pixel level variations follow recognizable patterns, and large scale texture where features, rather than pixels form patterns. Small-scale colour-texture analysis has been used as part of a segmentation scheme [14] and has also been used in the generation of synthetic test images [18]. Large-scale texture analysis has attracted more attention — such patterns are identifiable by humans as features. The surface of most areas of skin (except palms and soles) is covered in a network of fine lines that are a product of the structure of the top layer of the epidermis. Clinical features for differential diagnosis include disruption of the skin surface (erosion or crusting) and the presence of irregular clumps of abnormal cells in the upper dermis [1, 5]. These features can be seen in the disruption of the skin line patterning across the lesion, for example, the consensus statement of the USA National Institutes of Health [6] states that earliest melanoma can alter these skin markings. Consideration however of loss or accentuation of skin markings are controversial as indicators of early melanoma [19, 20]. Skin patterning has previously only been investigated in terms of changes in roughness of the skin surface topology perpendicular to the primary skin line

direction [20, 21] rather than changes in the pattern itself.

1.3 Thesis Outline

1.3.1 Aims

This work forms part of a programme which aims to provide a low cost image–based diagnosis support system. The programme will develop image processing techniques that will allow the formation of a vector of feature estimates relating to the identification of skin cancer from an image alone. Specifically, this project will:

- analyse the needs and role of computer based skin lesion analysis and to identify implications in terms of the system to be developed;
- improve the accessibility, for improvement and integration, of the 'edge focusing' boundary finding technique which was proposed and developed as part of previous research;
- investigate the possibilities for the development of metrics based on both colour and texture which would provide diagnostically useful information.

1.3.2 Structure

In order to appreciate the importance and challenge of pursuing automated diagnosis of skin cancer it is first necessary to discuss the high and rising incidence and the reasons suggesting that this will continue, as well as the nature of these cancers. Chapter 2 describes the nature and function of skin, details the most common forms of skin cancer showing why melanomas need special attention, briefly discusses their most significant cause and why this leads us to believe that the rate will continue to rise, and finally introduces the other features found on the skin which are most commonly confused with pigmented skin cancer.

Chapter 3 concentrates on the diagnosis of skin lesions and current methods used in the recognition of skin cancer. Existing medical guidelines for the identification of suspicious lesions are detailed together with their strengths, weaknesses and specific aims. Computer based diagnosis of skin lesions is introduced and discussed in terms of the need that such methods are intending to address. A review of research towards, and the current status of, such systems is made. Finally the aims and goals of a diagnosis support system for skin cancer are analysed and an outline description of the requirements and proposed structure for an effective system are synthesized.

Chapter 4 concerns the identification of lesion extent. The importance of this information in terms of directly obtainable diagnostic indicators and the enabling of both lesion area, and comparative lesion/surrounding skin evaluations is explained. The inherent problems and difficulties encountered in identifying the border are highlighted. A brief overview of research into boundary finding for skin lesion images is given as a prelude to the description of the *Edge Focusing* process which was the product of earlier research [16, 17]. In order to facilitate the further development, improvement, evaluation and integration of the edge focusing technique, the existing process was reconstructed in the Khoros II environment [22, 23]. The process and results of this porting into the highly modular and more intuitively accessible form within Khoros II are described and discussed. Finally, the limitations of the edge focusing technique are reviewed in relation to its use in the later sections of this study, the possibilities for improvements are discussed and the boundary definition policy employed in this work is presented.

Chapter 5 details the research undertaken into the possibilities for obtaining diagnostic feature information through the use of colour data. The importance of colour in diagnosis of skin lesions is widely recognized and reflected in the inclusion of colour features in both skin cancer checklists and in differential diagnosis descriptions. The concept of colour is introduced and the implications in terms of colour model resulting from requirements of the envisaged diagnosis-support system are discussed. The particular features pursued in image based diagnosis of skin lesions are identified together with the different methodologies used in their detection and computer based analysis. Current research into colour image analysis (the methodology most suited to the envisaged system) particularly for

skin lesion images and their segmentation is reviewed. The underlying goals of segmentation, particularly in relation to lesion images, are investigated in detail and conclusions are drawn resulting in the development of a new region-based technique. Initial results prompt a detailed consideration of other colour-spaces together with suitable colour similarity measures and patterns of lesion data distribution for them. Finally the segmentation performance on the new transformed data is presented and discussed.

Chapter 6 and 7 consider texture. Chapter 6 begins with an analysis of the possibilities for obtaining diagnostic feature information through the study of this data. The nature and concept of texture in image processing is discussed and the different analysis paradigms are analysed with reference to the type of texture they aim to model. The analysis of the large-scale texture of skin patterning forms the focus of the investigation; existing techniques are found to be inadequate for the description of this texture. A detailed investigation of the nature of skin patterning in lesion images is undertaken from which conclusions as to the requirements for modelling of this line segment based pattern are drawn. The abstracted representation is constructed in view of the need to capture the essential properties which would allow the measurement of disruption. A new technique is presented which is effective in extracting a representation of the quality of the skin line patterning.

Chapter 7 adresses feature analysis and evaluation using the skin pattern representation. Quantitative interpretation of the patterning information is vital if it is to be useful in an automated detection system. The provision of a metric for the quantification of disruption is therefore considered. Preliminary results for a number of analysis techniques prompt a number of changes and enhancements to both the extraction, analysis and evaluation methods. The final results show the effectiveness of this texture analysis and disruption feature evaluation in relation to both visual assessment and diagnostic performance.

Finally, chapter 8 draws general conclusions regarding the work presented and possibilities for the direction and specifics of future work are given.

1.3.3 Contributions

In the investigation of the role of computer systems in skin cancer diagnosis support and in the development of image processing techniques to provide quantification of diagnostically valuable indicators this thesis makes the following contributions:

- A review of the current skin cancer situation and an analysis of need for, and role of, computer based diagnosis support. This part of the study culminates in the synthesis of a set of aims and requirements for an effective system and a concept for the nature of such a system.
- The conversion of the edge focusing system for accurate and robust lesion boundary identification into a more modular and accessible form for future development and use. The evaluation of this system showing the success of the conversion in relation to these goals, and the price in terms of computational efficiency.
- An investigation into the role of colour analysis for skin lesions which results in both the identification of segmentation as the heart of such processing, and the fundamental properties of effective segmentation for work with lesion images.
- The development and implementation of a region based colour segmentation technique designed to cater for the identified need for locality consideration. The new technique differs from existing techniques used for lesion image segmentation as these focus on the classification of each pixel as a separate entity.
- A study of colour representations in relation to the identified need to reflect human judgment of colour consistency and an analysis of lesion image colour properties viewed in these spaces. The improvement in segmentation effectiveness when using the region based technique on a colour-space designed to be consistent with human perception.
- The identification of the large-scale texture of skin patterning and its disruption as a source of diagnostically useful information. The proposal for

an abstracted representation for this pattern which allows detection of its disruption.

- The development and implementation of a new technique for the extraction of skin patterning information from lesion images together with methods for analysing this pattern in terms of disruption which can be related to skin structure distortion by the lesion. The complete extraction, analysis and evaluation process is demonstrated to be effective in reflecting human appraisal of patterning disruption and in terms of diagnostic performance.

- Contributions to published literature.

  - A. J. Round, P. J. Fish, and A. W. G. Duller, "Khoros implementation of edge focusing system for skin lesion analysis," in *Proc. 18th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 1187–1189 (4.5.4–7), 1996.

  - A. J. Round, A. W. G. Duller, and P. J. Fish, "Lesion classification using skin line texture," in *Proc. Medical Image Understanding and Analysis*, pp. 169–172, University of Oxford, July 1997.

  - A. J. Round, A. W. G. Duller, and P. J. Fish, "Colour segmentation for lesion classification," in *Proc. 19th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 582–585 (2.3.4–3), Nov. 1997.

  - A. J. Round, A. W. G. Duller, and P. J. Fish, "Skin patterning analysis for lesion classification and segmentation," in *Proc. 9th Int. Conf. on Biomedical Engineering*, pp. 448–450, Singapore, Dec. 1997.

  - A. J. Round, A. W. G. Duller, and P. J. Fish, "A comparison of skin patterning feature analysis methods for lesion classification," in *Proc. SPIE* (K. M. Hanson, ed.), vol. 3338, Medical Imaging 1998: Image Processing (part 1), pp. 202–210, Feb. 1998.

Chapter 2

Background

In order to appreciate the importance and challenge of pursuing automated diagnosis of skin cancer it is first necessary to discuss the high and rising incidence and the reasons suggesting that this will continue, as well as the nature of these cancers. This chapter describes the nature and function of skin, details the most common forms of skin cancer showing why melanomas need special attention, briefly discusses their most significant cause and why this leads us to believe that the rate will continue to rise, and finally introduces the other features found on the skin which are most commonly confused with pigmented skin cancer.

2.1 Skin

The importance of the skin in the function of the human body is often underestimated, being viewed as simply the covering for the important organs. The truth is that the skin is one of the largest and most complex organs of the human body, performing an impressive range of functions and upon which we are absolutely dependent.

The skin has a typical surface area of $1.8m^2$ and accounts for some 16% of body weight [1]. It performs functions of sensing, vitamin D production, cooling and insulating and of course it forms the first line of defense against biological, chem-

ical, UV-radiation and physical damage. In order to perform such an array of functions the skin contains many cell types and this in turn means that there are many ways in which it can misfunction leading to a multiplicity of skin diseases [5].

Skin disease accounts for 10-15% of the general-practitioner's work in the UK and as such is one of the most common causes of loss of work [1, 5].

2.1.1 The Structure of Skin

Skin is composed of three layers: epidermis, dermis and subcutis (figure 2.1) the character and functions of each are outlined below [1, 5]. Exact structure and thicknesses are dependent upon location.

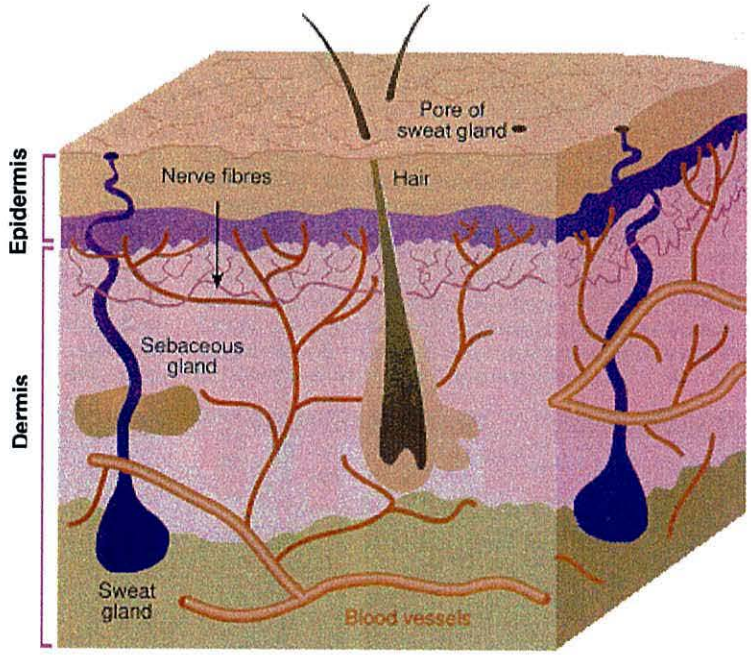


Figure 2.1: Diagram of the skin (adapted from New Scientist [24])

Epidermis. The epidermis is the outer-most part of the skin and perhaps the most complex in structure. Typically it is about 0.1mm thick (but may be up to 1.4mm on the palms and soles of the feet). The epidermis is a stratified structure

composed (from surface inwards) of the horny, granular cell, prickle cell and basal layers (figure 2.2).

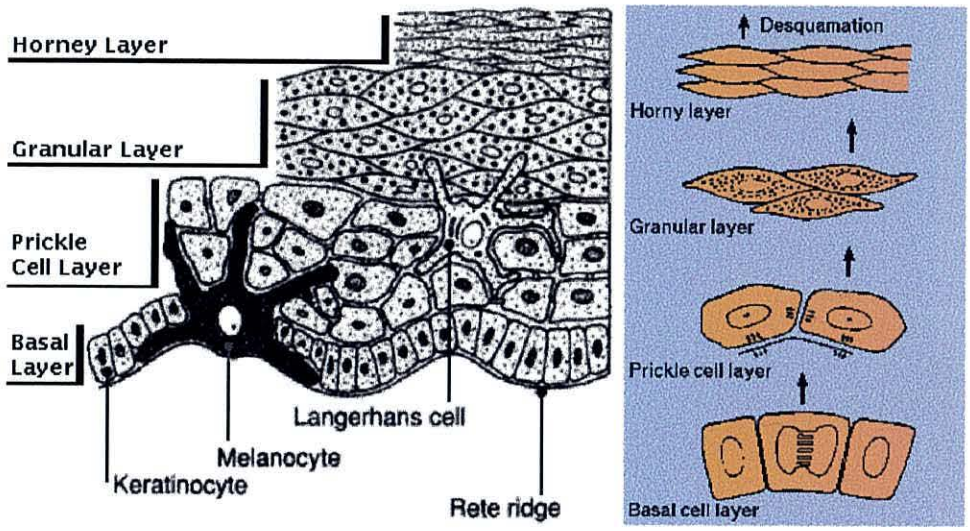


Figure 2.2: Diagram of the epidermis and the keratinocyte maturation process (adapted from Gawkrodger [1])

The epidermis is constantly undergoing a sequence of cell maturation which begins at the basal layer and ends with shedding (desquamation) from the horny layer. The basal layer (stratum basale) is mostly composed of Keratinocyte cells which are constantly dividing to provide new (squamous) cells for the other layers, and melanocytes (one per 10 to 20 basal cells) which synthesize melanin – a dark pigment which acts as a natural sun-screen. The prickle cell layer (stratum spinosum) contains polygonal rather than columnar keratinocyte cells strongly connected to distribute structural stresses, and Langerhans cells which are part of the immune system. In the granular layer (stratum granulosum) the keratinocyte cells become flattened and loose their nuclei. These cells are strengthened and bonded together to form the horny layer (stratum corneum) which is composed of overlapping sheets of dead, flattened cornified cells with no nuclei (corneocytes). The pattern of criss-crossing lines on the skin surface (figure 2.3) is formed by these overlapping sheets of horn cells. When the process that forms the horny layer (keratinization) is disturbed, the horn cells no longer shed normally and the skin patterning is disrupted. The horn cells are strongly bonded together and form an effective first line of defense barrier against mechanical, chemical, biological and UV-radiation damage.

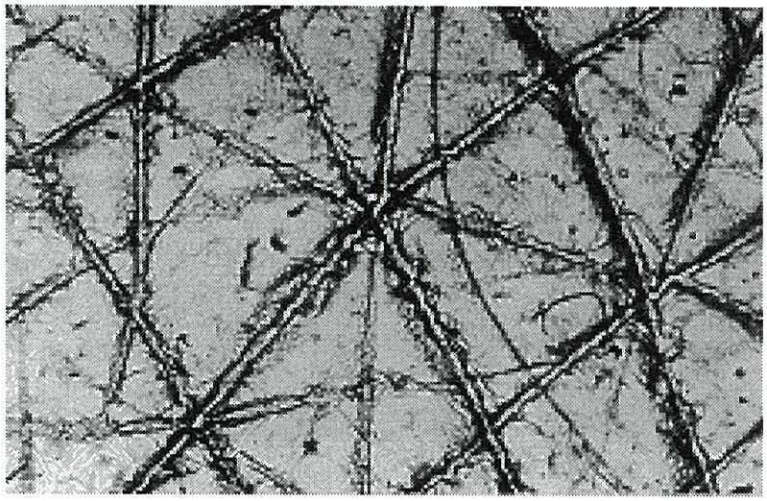


Figure 2.3: The surface of the epidermis (horny layer) showing skin patterning typical of the forearm. (adapted from Marks [5])

Dermis. The dermis is defined as a strong and elastic supportive matrix of connective tissue and contains specialized structures such as and concerned with hair follicles, sweat glands, blood vessels, immune cells and nerve fibres. The dermis is responsible for much of the skin's functions of mechanical protection and thermoregulation with the combination of sweat production and regulation of blood flow in the outer dermis.

Subcutis. The subcutis consists of loose connective tissue and fat which provides insulation.

2.2 Skin Cancer

The current incidence of skin cancer rivals that of all the other forms put together [1–3]. In the US there are about a million new cases of skin cancer annually [2, 3]. It is most common in light-skinned people exposed to intense sunlight [1, 2], white Australians have the highest rate of skin cancer in the world [2, 4] – only 0.3% of the world's population accounting for 6% cent of all the lethal forms of skin cancer diagnosed globally [4].

Skin cancer typically takes on one of three forms corresponding to the three major

types of cells in the epidermis: basal cells, squamous cells and melanocytes [1, 2].

2.2.1 Basal Cell Cancer (BCC)

Basal cell cancer[1] (figure 2.4) is the most common form [1, 2, 4, 5, 25] and accounts for about 80% of all skin cancer cases [3, 4]. By far the most common type is initially seen as small, pinkish raised area which grows slowly but relentlessly sometimes becoming darkly pigmented and eventually forming a central ulcerated region with an adherent crust [1, 5]. They occur most often in the elderly and middle-aged on skin that receives the greatest exposure to ultraviolet radiation - the face, ears and neck [1, 4, 5]. These cancers are aggressive and locally invasive, but rarely metastasize[2] and as a result a simple excision generally (95% [1, 3, 5]) effects a complete cure.

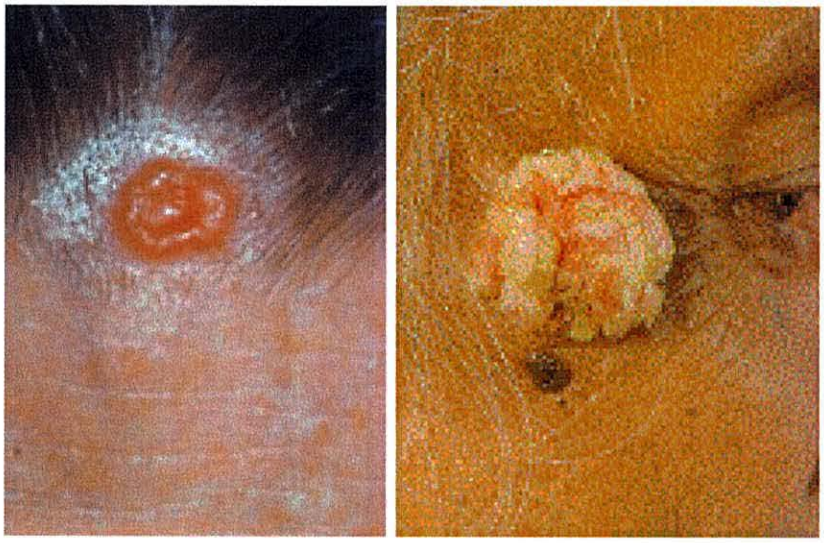


Figure 2.4: Left: Basal Cell Carcinoma. Right: Squamous Cell Carcinoma. (from Matrix/Loyola online [26] and http://biomed.nus.sg)

[1] basal cell cancer is also known as basal cell carcinoma, basalioma and basal cell epithelioma. carcinoma: malignant tumour of skin, epithelioma: skin cancer (from epithelium - epidermis)

[2] metastasize: movement of malignant cells to other parts of the body

2.2.2 Squamous Cell Cancer (SCC)

Squamous cell cancer[3] (figure 2.4) is the second most common skin cancer [4, 25] accounting for about 10 to 20 per cent of skin cancers [4]. They mostly appear as irregular dome-shaped nodules[4] which can ulcerate[5] and form a crust [1, 4, 5]. They typically affect the elderly and middle-aged, but unlike BCC an appreciable portion occur earlier. This cancer is most closely related to exposure to the sun [5, 25] and appears on the face, ears, the top of a bald head, and on the hands and arms [1, 4] often arising in areas of damaged skin [1, 26]. These are more dangerous than BCC; they grow faster and if allowed to develop, can metastasize, spread to surrounding lymph glands and be fatal [1, 4, 5]. Surgical excision is a successful cure in about 95% of cases [3, 5].

2.2.3 Malignant Melanoma (MM)

Malignant melanoma[6] is the most lethal of the skin cancers, accounting for only about 5% of cases but 80-85% of fatalities [2–4]. Incidence is increasing at 7% or more per annum [1, 3, 5] with British deaths doubling every 10 years [24]. These pigmented cancers have an irregular margin and often display a marked variation in colouring[7] together with erosion/crusting of the surface [5]. They can occur anywhere on the body but are most common on the trunk and lower legs [1, 4, 5, 8], often occur (about 50% [5]) on the site of a pre-existing benign melanocytic naevi, but can also spontaneously appear [1, 5, 8]. Solar UV-radiation is believed to be the single most important causative factor [4, 5, 27] but up to 50% occur in sites not normally exposed to the sun and although these may be attributed to occasional (holiday) intense exposure with sunburn [4, 5], other factors (such as developmental anomalies) may be involved [5]. If left untreated, a malignant melanoma is usually fatal [4].

[3] Squamous cell cancer is also known as squamous cell carcinoma/epithelioma

[4] nodule,papule: a raised area on the skin (not caused by free fluid or puss). Nodule when diameter is over 5mm, papule if smaller.

[5] ulcer: an area of skin loss extending through the epidermis and into the dermis

[6] Malignant melanoma is also referred to as just Melanoma

[7] variation in colouring is often termed variegated colouring

Malignant Melanomas have been subdivided by their patterns of growth and invasion, and hence their prognosis. The types, and their properties are: (figure 2.5)

Superficial Spreading Malignant Melanoma (SSM) is the most common form of melanoma, accounting for 50+% of cases in the UK [1, 7, 8, 26]. It affects ages 20-60 [1] but primarily 40-60 [8]. It is twice as likely in women [1, 8] where it is most common on the lower legs. In men it is most common on the trunk, especially the back. The tumour is macular[8], has an irregular outline which may be notched, and has pigmentation that is often haphazard with tan, brown, black and even pink and white in areas of regression. Early growth is radial and confined to the epidermis, later vertical invasion can occur and it is then that prognosis rapidly decreases. Lesions are unlikely to show substantial distortion of skin creases until there is significant tumour activity in the dermis [8].

Nodular Malignant Melanoma (NM) is the second most common form, accounting for 25% of cases in the UK [1, 8]. It is more common in middle age and in men. This form is often diagnosed quite late (perhaps because the documented characteristic features tend to be those of late stages in development – large blue-black nodule with ulceration and bleeding) [8]. Early lesions (sometimes called *Papular* rather than Nodular MM) are characterised by a small raised, darkly pigmented area which will continue to grow, sometimes rapidly. At first glance they appear regular, but closer examination shows notching or a streak of pigment at one edge. These lesions can sometimes be mostly de-pigmented but still usually show a rim of pigment [7, 8]. They quickly invade downward, often metastasize, and have generally poor prognosis.

Lentigo Malignant Melanoma (LMM) accounts for 15% of UK cases [1, 7] and most often occurs on sun-damaged skin on the face of persons aged 70+ [1, 8]. It has a slow radial growing pre-malignant phase referred to as Lentigo Maligna (LM) and Hutchinson's Melanotic Freckle or Hutchinson's Lentigo which can last up to 10 or 20 years [5, 7]. The LM stage starts as a tan-coloured macule resembling a stain which slowly evolves with dark brown and black and sometimes (due to regression) blue-grey and even white patches within the well defined but

[8] macule: a localised area of colour change without elevation or infiltration

highly irregular border of the lesion. If left untreated a true MM often develops within the LM which is then an LMM and has the characteristics described above [1, 5, 8].

Acral Lentiginous Malignant Melanoma (ALM) is the least common form in whites (accounting for about 7% [1, 8] of UK cases) but is the most common in black and oriental populations [1, 5, 8] who have a lower incidence of melanoma in general [5]. The tumour affects the palms soles and (rarely) nail beds, and can display any of the features of the other types [8] but is often diagnosed late [1, 8] when bigger, nodular and ulcerating [8] and hence has poor prognosis [1, 5].

Melanoma In Situ (MIS) has only relatively recently become a recognized melanoma type [6, 8]. These lesions are confined to the full thickness of the epidermis and do not invade vertically. Their appearance is similar to SSM being flat, asymmetrical lesions with notched, scalloped or jagged borders and a mottled brown pigmentation which can be tinged with blue, black or pink [8]. MIS can occur in any location [8] and can be treated effectively by conservative surgery [6].

2.3 UV-Radiation and the skin

Solar UV-radiation can cause all three main cancer forms in the skin: It is quite clear that most BCC and SCC are caused by chronic solar UV exposure [5], but perhaps more importantly, solar UV is believed to be the single most important causative factor [5, 6] in the most dangerous of these cancers, Malignant Melanoma.

UV radiation is part of the spectrum of EM radiation that penetrates the atmosphere (figure 2.6). It is divided into bands, A,B and C of decreasing wavelength. All three cause damage to the skin known as photoaging; photoaged skin is coarse, wrinkled, pale yellow and irregularly pigmented and more prone to biological attack and malignant lesion development [1]. UV-C has the most energy and is the most harmful, however very little currently reaches us because of absorption by atmospheric ozone. Some exposure to UV-B is important as it promotes the syn-

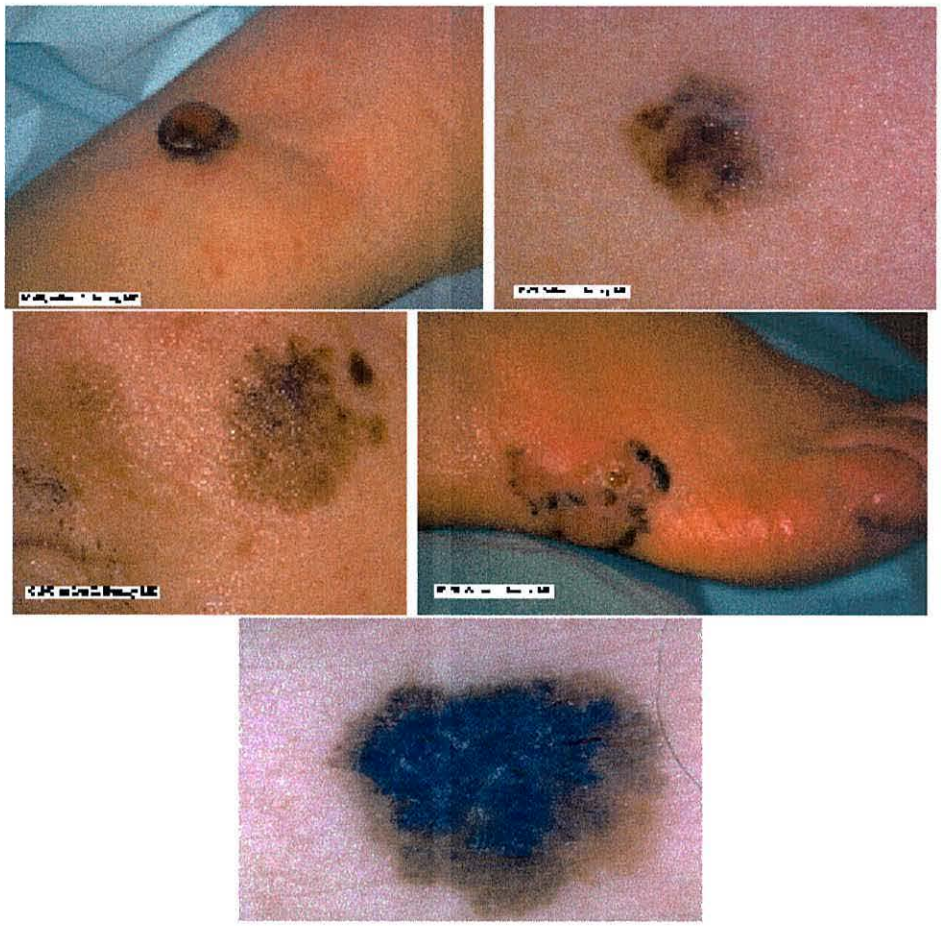


Figure 2.5: Malignant Melanomas: Top: Nodular MM and Superficial Spreading MM. Centre: Lentigo MM and Acral Lentiginous MM. (from Matrix/Loyola online [26]) Bottom: Melanoma in situ (from sample image set)

thesis of vitamin D_3 in the skin, however UV-B is much more damaging (about 1000 fold [5]) than UV-A. (even though most (90%) of UV-B is absorbed by the epidermis [1]). A *suntan* is the skin's shield against the sun's harmful rays. It is caused by both UV-A and UV-B and is mostly a UV-absorbing blanket of melanin in the horny layer of the epidermis [1].

One aspect of photoaging that is of particular importance in the study of skin cancer is the formation of *actinic(/solar/senile) keratosis (SK)*. It is a pre-malignant form of squamous cell cancer, and although transformation is rare[9] [1, 5], they indicate that chronic solar UV exposure has occurred and that cancerous lesions

[9] Although the majority of sources say that transformation is rare, it is also viewed as *slow* but not rare [28].

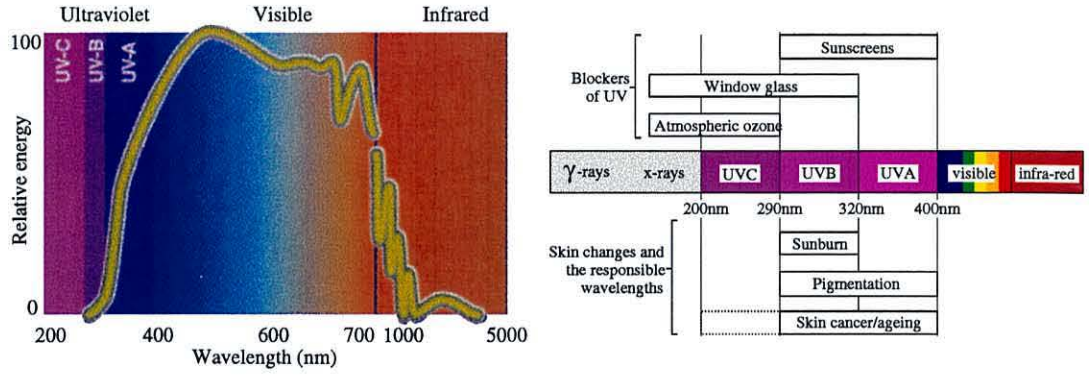


Figure 2.6: UV radiation reaching the earth's surface (adapted from New Scientist [24] and Gawkrodger [1])

are more likely to arise [5]. SK lesions are mostly multiple scaly or warty plaques or papules 2-5mm in diameter and are accompanied by other signs of sun damage. They occur on exposed sites and are most common in fair-complexioned elderly males [5] but can occur as young as 20, becoming progressively more common with age. In subtropical Australia SK is found in more than half of the population over 40 [5].

2.4 Benign Lesions

All malignant lesions offer improved prognosis if treated early, and this is especially true of Melanomas where a matter of 6 months can bring five-year survival prognosis from 95% down to 40% [7]. Unfortunately, early cancerous lesions (especially pigmented forms) often resemble benign lesions such as moles, features that are both extremely common, and can spontaneously arise and alter in appearance [1, 5, 7, 8].

The most common of the benign pigmented lesions necessary for differential diagnosis with cancer lesions are Melanocytic Naevus, Basal Cell Papilloma, Dermatofibroma and Vascular Malformation [1, 5, 7] (figure 2.7).

Melanocytic Naevus (mole) is very common, especially in whites who are likely to have several. They are generally small, have even brown colour and

smooth outlines. A naevus is a benign proliferation of the normal skin cells, most commonly the melanocytes [1]. The type of mole is often defined by depth and includes – *junctional* (epidermis/dermis junction) a flat light or dark brown, round or oval region 2-10mm diameter mostly on palms, soles or genitalia, *intradermal* (in dermis) a skin coloured or pigmented dome shaped papule or nodule mostly on the face and neck, and *compound*[10](both places) a smooth[11] papule usually less than 10mm diameter often tan or brown that can occur anywhere [1].

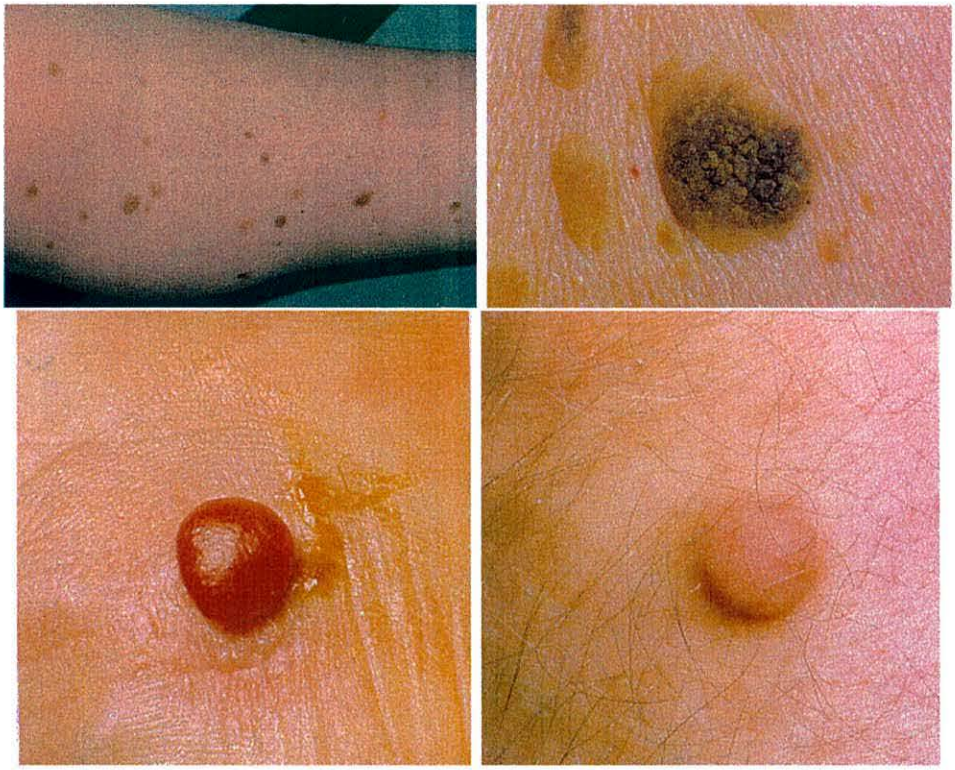


Figure 2.7: Benign Skin Lesions: Top: Benign Melanocytic Naevi (moles) and Basal Cell Papilloma. Bottom: A Vascular Malformation Lesion and Dermatofibroma. (moles from Gawkrodger [1] rest from Marks [5])

Dysplastic[12] naevi show unusual features of irregular outline, irregular pigmentation and size over 7mm diameter [1, 5], they should be carefully watched or removed and indicate an increased likelihood for development of melanoma [1, 5, 6, 8]. Moles can be present at birth (congenital) or can develop later, (acquired) mostly before the mid 20's. The congenital type is rare, affecting about 1% of

[10] Compound Naevus is often abbreviated as CN

[11] Small CN are generally smooth but large lesions can be warty

[12] Dysplastic, conveys a "funny looking" appearance and means having "heterogeneity of cell and nuclear size shape and staining" [5]

whites, however the lesion can be large, is sometimes disfiguring and can in some cases develop melanoma cancer later in life. The acquired type is much more common with the average white young adult having 20 or more [1, 7], and can be mistaken for the first appearance of melanoma.

Freckles and Lentigines are not true naevi, but share many characteristics. They are both small light brown macular regions, but freckles are regions with more melanin which darken in the sun while lentigines have more melanocytes rather than melanin and do not darken [1].

Basal Cell Papilloma (BCP)[13] [1, 5, 7] is an extremely common lesion in aging skin (most people over 40 have a couple, some have hundreds[5]). They start as a small papule, often light tan or yellow and will grow becoming a dark brown and warty nodule 1-6cm diameter with a "stuck on" appearance and well defined edges[1]. They are often (but not always) multiple and appear on the trunk (particularly on normally covered skin[7]), neck and face. They are totally benign[7], but dark lesions can resemble melanomas.

Vascular Malformation[14] [1, 5] covers a multitude of different lesions arising from the skin's blood vessels and capillaries. The majority of these lesions are not serious and many will heal quickly without intervention. Some formations resemble early melanomas, such as Capillary Aneurysm which suddenly appears as small black region and Pyogenic Granuloma, which develops in about a week but fades in about a month and is seen as a glazed or blood crusted red papule which can be mistaken for nodular MM. Vascular lesions are generally more red than brown and unlike melanocytic lesions they will blanche with pressure (unless thrombosed).

Dermatofibroma[15] [1, 5] is a common dermal nodule 5-10mm in diameter with a rough or warty surface due to thickening of the epidermis above the lesion. They are most common on the lower legs of young adults, especially women.

[13] BCP is also known as seborrhoeic wart or seborrhoeic keratosis ($\neq SK$!)

[14] Vascular malformations are also referred to as vascular naevi, angiomas and haemangiomas and include vascular dilation lesions such as stork mark,port wine stain and venous lake, and vascular proliferations such as strawberry mark, senile angioma(Campbell-de-Morgan spot), cherry angioma and other capillary angiomas

[15] Dermatofibroma is also known as Histiocytoma and Sclerosing Haemangioma

They enlarge slowly if at all but are generally pigmented and so resemble early melanoma, however they are of no serious clinical significance themselves.

2.5 Conclusions

The incidence of skin cancer is high and rising across the world, a trend which is likely to continue given the importance of solar UV as a cause and the modern propensity to indoor lifestyle with sun-seeking holidays.

Malignant Melanoma is the most deadly form due to its rapid development, invasion and metastasis cycle. Early diagnosis is paramount as prognosis is hugely improved where the tumour is excised quickly. Many common benign pigmented skin lesions can resemble early melanomas which means that diagnosis *in situ* is both necessary, as it is not practical to excise all such lesions, and difficult as the early differential signs are hard to detect without detailed expert inspection.

Chapter 3

Diagnosis & System Concept

This chapter concentrates on the diagnosis of skin lesions and current methods used in the recognition of skin cancer. Existing medical guidelines for the identification of suspicious lesions are detailed together with their strengths, weaknesses and specific aims. Computer based diagnosis of skin lesions is introduced and discussed in terms of the need that such methods are intending to address. A review of research towards, and the current status of, such systems is made. Finally the aims and goals of a diagnosis support system for skin cancer are analysed and an outline description of the requirements and proposed structure for an effective system are synthesised.

3.1 Diagnosis

Medical diagnosis of skin lesions involves the use of two distinct classes of examination, invasive and non-invasive. Histology[1] is often regarded as the only truly reliable method for establishing the nature of a lesion [6], however it involves the examination of samples taken from the lesion or the removal of the whole lesion [9]. Such invasive forms of identification are obviously unsuitable for routine identification due to the time, cost and inconvenience to the patient.

[1] Histology is the study of organic tissue. Histopathology is study of change in skin due to disease.

Non-invasive examination is traditionally carried out by simple visual inspection although other methods such as ultrasound [29] and optical spectroscopy [30] have been applied.

Although simplest in application, visual inspection is not always conclusive — experienced dermatologists are generally around 70% accurate in the clinical diagnosis of non-typical pigmented lesions [5, 9] (although for experts with over 10 years experience accuracy is reported to reach 80% [95]). The complexity of the skin and the large number of cell and tissue types it contains results in an enormous number of lesion types. Effective differential diagnosis is difficult as these lesions present only a limited number of clinical appearances and often share visual characteristic features [5].

Diagnosis of skin lesions is further complicated by the desire for *early* diagnosis. Malignant lesions, especially invasive melanomas have a much increased fatality rate as the lesions develop [1, 5–7] and hence malignant lesions needs to be identified as early as possible in their development. Early detection and diagnosis is probably the most critical factor accounting for increasing melanoma patient survival rates [31]. It is important then, that diagnosis from *early indicators* is taken as a key consideration for the development of any diagnostic guidelines or system as the features commonly used in the description of malignant lesion types are often those of later development – new and clever methods are not needed to identify large ulcerating or bleeding lesions.

3.2 Current Medical Guidelines

Perhaps the most significant recent advance in the treatment of skin cancer has been the understanding that prompt excision of early tumours such as melanomas can effect a complete cure [8]. A large portion of recent efforts in this area have consequently been directed to improving the accuracy of initial diagnosis and speed of referral of suspect lesions to a specialist. There are two main obstacles to this aim: the patient must first be aware of the signs to look for and the GP (likely first point of contact) must be well informed or assisted in diagnosis (it is unlikely that a family doctor will encounter more than one melanoma in 10

years [8]). Information campaigns have been used throughout the world with the aims of increasing awareness of the risks of exposure to sunlight, encouragement of the public to self-examine and details for a variety of audiences of the features to look for.

Checklists have been proposed to aid in diagnosis by highlighting the most important indicators. MacKie [7] originally proposed a seven point list below which was adopted by the Cancer Research Campaign.

1. **Itch**	Itching or other change in sensation.
2. **Size**	Length of largest diameter greater than 1cm.
3. **Increasing Size**	Growth of a lesion.
4. **Shape**	Irregular outline, notch in border.
5. **Colour Variation**	Speckling and other variation in colours, esp. red, and blue-white.
6. **Inflammation**	Inflammation at the edge of the lesion.
7. **Crusting or Bleeding**	Slight oozing causing crusting.

The presence of three of the features is considered suspicious and if four or more are found the lesion is cancerous in over 90% of cases [7]. Emphasis on change in a lesion was introduced when the list was revised as this was found to increase the usefulness of the checklist as well as make it more memorable [32]. The presence of one or more of the major signs warrants referral and one or more minor signs suggests further consideration.

Major signs	**Minor signs**
change in size	inflammation
change in shape	crusting or bleeding
change in colour	sensory change
	diameter 7mm or more

Other checklists have also been proposed such as the ABCDE system of the American Cancer Society.

Asymmetry	Two halves of a lesion do not match.
Border Irregularity	Edges are ragged, notched or blurred.
Colour	Pigmentation is non-uniform. Shades of tan, brown and black with dashes of red, white and blue adding to the mottled appearance.
Diameter	Greater than 6mm and growing.
Elevation	Elevated by 2mm or more compared to surrounding skin.

This system is widely quoted in public information sources (such as the web pages of the American Academy of Dermatology and Skin and Cancer Foundation of Australia) but without the final Elevation feature. The "ABCD's of melanoma" is also advocated in the USA National Institutes of Health (NIH) consensus statement on melanoma [6] which adds that earliest melanoma lesions are flat or macular and may have altered skin markings, but that rapid change in any otherwise benign-appearing macular or palpable lesion can represent melanoma.

Although these checklists have been successful in many respects it has been noted that not only do they tend to be biased toward the detection of Superficial spreading melanoma and less sensitive to the biologically more aggressive nodular form [8], but that clinicians are inconsistent and have great difficulty in agreeing about the presence or absence of lesion features [33].

Activities such as the many information campaigns warning the public about the dangers of excessive exposure to sunlight and illustrating the signs that indicate a cancerous lesion are perhaps the only effective measure that can be taken at the population level. Although the advantages of earlier detection in many ways make melanoma an ideal candidate for screening [6, 8], the role of generalised screening is controversial [10] due to the prohibitive scale, cost and complexity of any effective programme. The whole body would have to be examined as malignant lesions can occur on any site, the average person has a large number of benign skin lesions each of which would have to be assessed and the screening would have to be very regular and comparative to previous data as change is a vital indicator and serious development in melanoma can occur in a few months. Such problems with screening activities have been found in free cancer screening clinics held at outdoor social events in the USA where the published results

suggest that they are labour intensive and expensive exercises for their yield [9]. It is also worth noting that such events are not true *screening* activities as it is reasonable to suppose that the people who ask for examination regard the mole(s) they have as suspicious and so screening of a population is likely to have an even lower yield.

Encouraging self referral is much more practical, where personal responsibility is taken for detection of change or other suspicious features. This does however require that such referrals can be dealt with both quickly and accurately and with minimal inconvenience to the patient. In addition, surveillance activities are sensible (and advised) on a strictly limited population deemed to be high-risk, such as patients with dysplastic naevi or a personal or family history of melanoma. Even in these cases attempts to detect change in lesions often cause problems of information overload in terms of both storage and management [29].

Whatever the debate concerning the most effective form for public or GP information it is generally accepted that the most common signs of early melanoma are increase in size, change in colour or shape and itching or increased awareness [31]. Hall [8] provides a good summary of the difficulties involved with melanoma identification and the general signs to be looked for by saying:

> "The majority of us acquire new moles through childhood into early adult life. These lesion will grow in diameter and height, become darker or lighter and may even regress and disappear by the time we are in our forties (Fitzsimmons 1984). Any new mole arising in the late twenties and early thirties and any mole growing out of step compared with other pre-existing naevi at any age including childhood should arouse suspicion. sudden new growth in a previously dormant mole clearly must be regarded as abnormal. Disorder in growth, border or pattern of pigmentation are the things to watch for (Sober 1985)."

Tumour thickness is widely regarded as the single best indicator for prognosis [1, 5–7, 34]. The standard measure, Breslow thickness, is defined as the distance in millimeters between the nominal skin surface and the deepest part of the tumour [34]. Invasion is also measured in terms of the Clark staging method. The five stages correspond to the stratified structure of the skin with a stage I tumour confined to the epidermis, stage V indicating infiltration of the subcutaneous fat

and stages II-IV reflecting progress through the dermis. These measurements are normally obtained by microscopic analysis of stained lesion section samples. Attempts have been made to estimate the tumour thickness non-invasively from high-frequency ultrasound images [8, 29], however only Clark stages IV and V can be reliably detected and ultrasound based thicknesses can differ considerably from histologically derived values [29].

3.3 Computer Based Diagnosis

Placing The Need

The gains in terms of improved prognosis where a malignant lesion is treated early in its development demand that every effort is made to reduce the time taken in the process of discovery, first diagnosis, expert diagnosis and intervention. Improvements in the initial discovery phase beyond campaigns of public information for self examination have been shown to have many practical difficulties whilst surgical intervention by excision is both quick and effective as a treatment. This suggests that improvements are most likely and valuable in the diagnosis chain.

It is widely recognized that differential diagnosis of pigmented lesions is a difficult enterprise even for specialist dermatologists. The role of the GP demands an extremely broad expertise covering the whole range of ailments that are presented at a local surgery. The heavy requirement for specific depth in the diagnosis of skin lesions is therefore not one that the family doctor can be expected to meet and yet they must deal with the cases presented to them. The accuracy of the family doctor is unlikely match that of the dermatologist [9, 10] (although no study of this has yet been published perhaps due to the difficulty of ensuring statistical reliability). It would seem then, that the most effective way of improving the efficiency in the diagnosis chain would be to target the initial diagnosis in the local surgery, and that a diagnosis support system could aid the GP by embodying the specific details of pigmented lesion diagnosis which the breadth of their discipline denies them. The increase in self referral due to public information efforts could then be serviced more effectively by an improvement in the speed and accuracy of the initial diagnosis and therefore a reduction in the load on expert dermatologists.

Computer Based

Computer based analysis and diagnosis may well provide an avenue for the improvement of diagnostic accuracy and the speed with which it can be achieved. Computers offer objectivity and can provide consistent quantitative results, they can also offer benefits in storing records and cross-referencing as well as possibilities for enhancing and combining multiple sensing modes such as UV, IR and ultrasound as well as visual images. It has been noted, for example, that computers are ideally suited to the tasks involved in dysplastic mole surveillance – change detection, quantitative measurement and deduction [29]. There has been considerable effort toward the development of systems to perform differential diagnosis of melanoma and other pigmented lesions including artificial intelligence and neural networks for diagnosis (e.g. [35, 36]) and work directly on colour (e.g. [13, 37, 38]), shape and edge-finding (e.g. [15, 17]). Such systems often reach a comparable accuracy to that of the dermatologists on their test data sets, however diagnosis is often restricted to a binary separation of two specific lesion types due to the complexity and number of possible diagnoses when dealing with the full range of pigmented lesions.

Image Based

It is possible for computers to be used only in the encapsulation of the knowledge for differentiation, with an operator supplying measurements and observations as prompted and the computer making decisions based on this information (e.g. AI/DERM [29, 39, 40]). However, this mode of operation does not utilize the full potential of the computer to provide a valuable service on several counts: the system requires a trained operator with understanding of how to quantify observations such as variegated colour, the complete process may take a considerable amount of dedicated time, the objectivity and consistency of any quantitative results could be questioned (it depends on operator skill) and finally the diagnosis cannot be viewed as truly independent where the GP is the operator.

The use of electronic digital imaging allows the required diagnostic measurements to be obtained directly by the computer from raw data. The operator no longer needs to be highly skilled or trained beyond proper use of the acquisition tool and the results can be viewed as an true independent analysis. The process would also take less time and involve less inconvenience to the patient as the need for

many measurements is avoided. In addition the advantages of data storage, cross-referencing and change detection are enhanced as new measurements from existing raw image data can be taken as required without the necessity of re-examination of the patient.

Visible Light Imaging

Electronic imaging offers the possibility of multiple imaging modes beyond that of visible light. Ultrasound, MRI, X-rays, UV and IR have all been used in dermatological analysis [29] with high-resolution ultrasound providing useful information especially on tumour thickness [8]. A combination of imaging modes may well prove essential for sufficiently accurate diagnosis, however much of the medical information on differential diagnosis of pigmented skin lesions relies on visual inspection. This is understandable since the skin is visible to the naked eye and is illustrated by the high proportion of the checklist indicators that are based on visible features. A consequence of the emphasis on visual inspection is that a large proportion of the features needed for differentiation can be obtained from a simple visible light image, features such as size, shape, border irregularity and asymmetry and variegated colour can all be directly obtained.

A system based on digitally captured visible light imaging has many advantages in the context of diagnosis support based in a local surgery. It satisfies the requirements of low capital and running expense; The home computer market has provided low cost video capture devices as well as personal computers of considerable power and the "cost" of an image is only seen in terms of disk space use. It has been shown that digitized images are no less informative in terms of diagnosis than are slides [41], which are often the current image storage medium and which are comparatively expensive, bulky and fragile. Digital capture time is close to instantaneous and patient inconvenience is minimal with visible light imaging, as there is no need for preparation or contact (except perhaps a bar to fix the range and provide a scale and colour matching chart). Use of the system would require very little training, especially with a fixed-focus and exposure. In addition (providing accurate calibration and known frequency rendering are not required) effectiveness of the equipment and acquisition can be checked without expert knowledge or training by simply checking for a good visual reproduction.

Recently there has been a great deal of interest in Epiluminescence Microscopy (ELM). This technique also uses visible light for imaging but the skin surface is rendered translucent by applying immersion oil and then covering the area with a glass slide. A new set of features is revealed by this process and the specular reflections which often hamper image analysis are avoided. Unfortunately the image quality is often reduced by the presence of small air bubbles in the oil particularly where the lesion is elevated or hairs are present. The new features revealed are complex – there is a significant learning curve involved in interpretation of the image observed [8]. This is reflected in evidence that, without training, the use of ELM images actually reduces diagnostic performance [10, 42]. In addition, the existence of *any* diagnostic advantage over a normal clinical images is controversial [32, 42].

A commercial hand held ELM imaging device called the dermatoscope is available and is being evaluated by many dermatologists and two systems, the Skin Polarprobe [43] and MoleMax II [44], use digital ELM images. Another instrument, the Nevoscope [45], has been developed which can provide the equivalent of ELM images without oil by using trans-illumination. This commercial device also provides standard images and several side and angle views and has associated image processing software [46] to calculate other features, however it has a small field of view and is bulky and expensive.

The disadvantages of ELM in relation to the identified diagnosis support system for GPs lie in the complication of the image acquisition, increase in patient inconvenience (albeit small), need for more training (especially if the quality of the acquisition is to be verified), and the possibility that well known medical diagnostic features will be altered or lost. Development of a system based on standard visible light images is therefore indicated and pursued here.

Vector of Independent Diagnostic Features

Design of systems based on a set of diagnostic features is not a new concept – the majority of current systems are based on indicators similar to those in the checklists above, although other features such as textural roughness [20] have been used. The formation of the set of estimates is however rarely promoted as a goal, the emphasis generally lying in the performance of evaluations from new

or improved quantification methods for one or more of a small set of particularly powerful or well known features. The performance is most often gauged in terms of discrimination of a set of two lesion types (often melanoma and benign naevus), an approach which is understandable given the complexity of the diagnosis situation, the many types of lesion and the desire to assess performance in terms clearly related to the system domain of application.

This situation has several negative impacts on progress toward a truly effective solution. Combination of features before evaluation precludes comparison with later work except in terms of effectiveness as a complete system or when considering the same set of indicators. The need for valid comparison drives the repeated investigation of the same indicators (e.g. size, border irregularity and variegated colouring). The desire for favourable comparison then results in particularly concentrated attention being given to indicators known to be particularly effective – starting with a less powerful indicator will reduce performance where this is measured as direct discriminative power and hence encouraging results are less likely. Finally, evaluation based on discrimination of a binary set of examples can be unhelpful or misleading in terms of the effectiveness of *feature* quantification as a certain feature may not divide the set exactly by lesion type any yet still be of considerable value in practice, for example surface disruption (e.g. crusting) seen in melanoma is also found in dysplastic yet nonetheless benign naevi.

The alternative approach of constructing piece by piece a vector of independent features offers many advantages and it would seem that every effort should be made to structure current and future work in this way. Under this paradigm the problem of diagnosis becomes separated from identification and development of feature estimators – there are many standard classification and reasoning methods and investigation of the relative merits of each is a large enough task in itself. Research into new features and the evaluation of improvements on existing features is also facilitated by this split and by the independent study of each feature – each element can be more thoroughly investigated, tested and evaluated and features need not provide a clean binary discrimination of medically defined types. Finally, the latest features can be more simply linked into existing final diagnosis implementations where these have been constructed with an abstracted and separate view of the feature vector and its compilation.

When constructing a vector of feature estimated, a number of considerations should be borne in mind:

- A global aim of high sensitivity must be maintained; it is essential that the final system keep to an absolute minimum the proportion of malignant lesions classified as benign – even if this is at the expense of some specificity, i.e. even if this results in more benign lesions being classified as malignant. This suggests the use of a larger number of features and the importance of a level of independence between them to avoid a single unusual aspect in a given case affecting all the estimates. The importance of multiple acquisitions via independent avenues of vital features is also highlighted, an example being boundary confirmation through comparison of texture, colour and intensity based methods as size, symmetry and boundary irregularity as well in-lesion (colour presence and variegation) and differential (lesion–surrounding skin) measurements and all rely on this information.

- The focus should be on distinguishing features which are effective on early lesions – analysis of very large, ulcerated and bleeding lesions should be de-emphasised and these should perhaps not appear in test sets for assessment of indicators (such lesions would probably warrant specialist attention regardless of diagnosis).

- It is desirable that the reasoning behind a diagnosis be humanly accessible – indicators should reflect a known medial indicator or at least have visual meaning wherever possible. In this way confidence in the diagnosis can be increased by human verification that the features do indeed appear and, of course, it allows the diagnosis to be explained to a patient.

- It is desirable that diagnosis can be presented as a case of recognizable features (related to the above point). Medical texts [1, 5] often approach differential diagnosis by proposing likely alternatives – a case-based best match approach. Transfer of knowledge from texts and experts would be facilitated with such a system and the diagnosis is made more accessible to professionals by reflecting standard medical practice. In addition case development can be compared over time allowing new developmental analysis.

The checklists provide an obvious first source of indicators and, as mentioned before, much research and many current systems are essentially based upon a subset of these. Direct utilization of just the checklist features for diagnosis is however questionable since they have been specifically designed with the overriding aims of both simplicity and of detecting all malignancies regardless of cost in terms of specificity. However, given the known effectiveness of these indicators and the importance of both explainable diagnosis and high sensitivity in the primary care situation of the envisaged system, these indicators should obviously should form a valuable part of a more complete feature list.

3.4 A Computer-Based System

From the considerations discussed above an outline description of an effective diagnosis support system for pigmented skin lesions has been synthesised:

Aim: To provide differential diagnosis support for pigmented skin lesions in the primary care setting.

Requirements:
1. minimum capital and operational cost
2. minimum patient inconvenience (also time)
3. minimum training for operation
4. robust and explainable diagnosis

From this basis conclusions have been drawn concerning the nature of a system best suited to the identified need and the structure of research towards this system.

	Conclusion	**drawn from**
i.	Based on normal visible light imaging.	(1,2,3,4)
ii.	Using Off-the-shelf technology	(1,i)
iii.	Tolerance to non-ideal capture	(1,3,4,ii)
iv.	High sensitivity to malignancy and good specificity	(4)
v.	Vector of independent feature estimates	(4,iii,iv)
vi.	Using visually verifiable features	(4)

3.5 Conclusions

The task of providing diagnosis support to aid in the fight against skin cancers is far from simple. The discrimination task itself is widely recognized as difficult even for experts in the field – no simple set of non-invasively obtainable information is known which allows even a near perfect diagnosis. Information campaigns for the public and GP's have presented many valuable indicators and several systems to aid in diagnosis.

Skin cancer, particularly melanoma shows hugely improved prognosis with relatively simple surgical intervention when the lesion is removed early in its developmental cycle, before invasion and metastasis. Earlier diagnosis and quicker referral times are therefore a priority and this suggests a focus on improving the speed and accuracy of primary care diagnosis. The breadth of knowledge demanded of a GP precludes the specific depth required given the difficulty and highly skilled nature of the discrimination task and this suggests the need for a diagnosis support tool.

An effective diagnosis support tool for the primary care sector must satisfy the key requirements of minimum capital cost, minimum training for operation, and minimum operating time, cost and fuss. The diagnosis needs to be highly sensitive so as to avoid missing any cancers whilst retaining a satisfactory specificity to avoid excessive referrals.

An image based system avoiding specific requirements on image capture and using simple off-the-shelf and therefore low cost home PC technology is indicated. The complex nature of the discrimination task suggests the development of a wide range of mutually independent feature estimates. These should be related to known medical indicators or visible features in order that confidence in the diagnosis can be boosted through interrogation of the "reasoning" underlying the final diagnosis.

Chapter 4

Boundary Finding

This chapter concerns the identification of lesion extent. The importance of this information in terms of directly obtainable diagnostic indicators and the enabling of both lesion area, and comparative lesion/surrounding skin evaluations is explained. The inherent problems and difficulties encountered in identifying the border are highlighted. A brief overview of research into boundary finding for skin lesion images is given as a prelude to the description of the *Edge Focusing* process which was the product of earlier research [16, 17]. In order to facilitate the further development, improvement, evaluation and integration of the edge focusing technique, the existing process was reconstructed in the Khoros II environment [22, 23]. In Khoros, the entire process is decomposed and presented in a highly visual manner with the basic processing actions linked by data flow lines. This presentation is ideal for experimentation since elements can be readily exchanged, allowing for simple evaluation of alternatives and for reuse of elements in future projects. The process and results of this porting into the highly modular and more intuitively accessible form within Khoros II are described and discussed. Finally, the limitations of the edge focusing technique are reviewed in relation to its use in the later sections of this study, the possibilities for improvements are discussed and the boundary definition policy employed in this work is presented.

4.1 Lesion Boundaries

Many of the important diagnostic indicators identified for the clinical assessment of malignancy in melanoma can be obtained from boundary information alone. Well known indicators in this category include size, border irregularity, notching and asymmetry [6, 7, 32]. The extent of the lesion is also a vital prerequisite for analysis of other features of the lesion such as texture and colour indicators [47], allowing for example the quantification of variegated colouring *in the lesion* and texture difference *between the lesion and surrounding skin.*

The identification of the lesion extent is however far from trivial. In order to retain diagnostic accuracy of features that directly or indirectly rely upon it, the acquired border must be of dependable accuracy, reliability and consistency. This means that the boundary must be correctly identified and reflect a consistent attention to detail in a manner robust to a wide range of image imperfections. This is perhaps especially important given the diagnosis support system as envisaged and its requirement to work with low cost equipment and minimal operator training. Images may include features such as rulers, image calibration scales, skin hairs and the edges of limbs, and the boundary needs to be identified in spite of these as well variations in lighting and camera position and orientation [16].

In addition to all these complications there is also considerable difficulty in providing a good *definition* of the lesion boundary: A poorly defined or 'blurred' edge is an indicator associated with malignancy as is inflammation — where should the boundary lie when the lesion fades into the skin, and should the lesion be taken to include a surrounding inflamed area? Specific types of lesion such as halo naevi (is the halo part of the lesion?) also present problems as do areas of regression in tumours (especially where these are adjacent to the surrounding skin) and tumours which are only part pigmented. In most cases the 'gold standard' used in analysis of performance relates to either outlines drawn by dermatologists or to simple visual inspection.

The importance in terms of diagnosis of, and the difficulties involved in obtaining, accurate and reliable boundary information has naturally been reflected in considerable research interest in this topic. Some research directly uses boundaries

drawn by an operator [48, 49] however this approach is both time consuming and heavily subjective. Automatic detection would be much more valuable for use in a diagnosis support system. A wide variety of automatic methods have been proposed, based on grey-scale intensity, colour or texture data. Multi-channel techniques are also used, merging information from more than one of these sources, and it is common to begin with colour data and pre-process this to provide a more useful single element data image than the simple grey-scale intensity, intending to capture the essence of the colour information.

Claridge *et al* [11] used a semi-automatic thresholding technique on grey-scale intensity images with operator confirmation of suitability being required. A threshold approach has many problems with the less distinct cases and may result in the identified lesion area being composed of multiple regions or containing holes where the threshold is too high, or in non-lesion areas being identified as lesion if the threshold is too low. Often there is no threshold which provides a truly successful result although the application of binary morphological closing can rectify some of the problems with holes and breaks caused by higher thresholds.

Umbaugh *et al* [12] and Ercal *et al* [13] present more sophisticated techniques based on colour data. These techniques transform the original colour-space and then use colour quantization and thresholding with region growing respectively to identify the lesion extent. Hance *et al* [37] compared a wide range of border identification techniques and conclude that of those tested, that of Umbaugh *et al* [12] performed best. Dhawan *et al* [14] propose a complex algorithm combining a pyramid segmentation scheme using intensity data, and co-occurance matrix evaluation of colour texture information. All these techniques and the review are considered in more detail in the chapter on colour analysis, and the texture element of Dhawan *et al* [14] in the chapter on texture.

Segmentation can be pursued not only by identification of homogeneous region extent, but also by finding edges (points of rapid change in image value) directly. Such techniques generally use a single element image data (such as grey-scale intensity) to avoid the complex issues of inter-element interactions. By looking for *change* across the lesion boundary problems of sensitivity to overall intensity variations are reduced. Perednia *et al* [50] used a fixed size Laplacian of Gaussian

(LoG) filter (LoG filters are described later in this chapter) on an image taken at low resolution in order to avoid unwanted small image features. Interpolation was used to find the lesion to sub-pixel accuracy. Golston *et al* [15] present an alternative method which uses a 'radial search' concept on a intensity image to find a boundary as part of a multi-channel approach combining colour, luminance and texture based boundaries in relation to confidence measures. The radial search algorithm considers data on radial lines extending from the image centre, which is assumed to be the approximate centroid of the lesion. The lesion border is detected as a sustained jump in grey level on these radial lines. The identified boundary points (excluding those significantly different to their neighbours) are joined using a B-spline curve. The method is computationally inexpensive but it requires that the lesion boundary cross each radial line only once.

Although not for lesion images specifically, snakes (also called active contours) [51] have often been proposed for the segmentation of medical images. The following are some recently published examples. Yezzi *et al* [52] propose a geometric snake method for segmentation of MRI, CT and ultrasound medical imagery. The method uses feature-based metrics and unifies curve evolution and classical energy minimization approaches and its effectiveness on relatively noisy images is illustrated for examples in each of the three imaging modalities mentioned above. Mikić *et al* [53] focus on dynamic segmentation for image sequences. They review some of the existing snake models before describing their new model which introduces optical flow estimates into the snake technique to aid in tracking moving boundaries in the image sequence. The technique is used to track cardiac structures in ultrasound image sequences and found to be generally successful even where large between-frame displacements occur. Although snakes are often effective, the success of these methods depends upon the choice of an energy equation which clearly reflects the boundary to be found. The difficulties in obtaining an exact definition of what constitutes a lesion boundary, as discussed earlier in this section, therefore significantly hamper the use of snakes for this application.

4.2 Edge Focusing

The *Edge Focusing* technique for lesion boundary identification forms a major part of previous PhD research [16] concerning medical image analysis at the University of Wales, Bangor. Edge focusing provides a means by which a detailed boundary can be obtained which is robust to many forms of contaminating image feature and image capture imperfections. In essence the technique, which is based on that proposed by Bergholm [54], works by selecting an optimal border from a sequence of increasingly fine scale estimates for the boundary of the most lesion-like object in the image.

The complete edge focusing system can be broadly divided into three stages; pre-processing, generation of an edge estimate sequence, and optimal edge selection.

Preprocessing

The goal of the preprocessor is to provide the approximate size and location of the most lesion-like object in the image. This is achieved by the sequence of operations illustrated in figure 4.1. The stages involved are: Smoothing with a 7×7 window median filter to remove small scale noise features, thresholding using the Kittler [55] process which automatically selects an 'ideal' threshold, in this case based on pixels of high gradient, morphological closing to remove very small regions and holes in the 'lesion' areas, tracing to convert the binary image into coordinate sets representing the separate objects identified in the image, and finally identification of the most lesion-like object by the application of a set of heuristics covering size, shape and location properties.

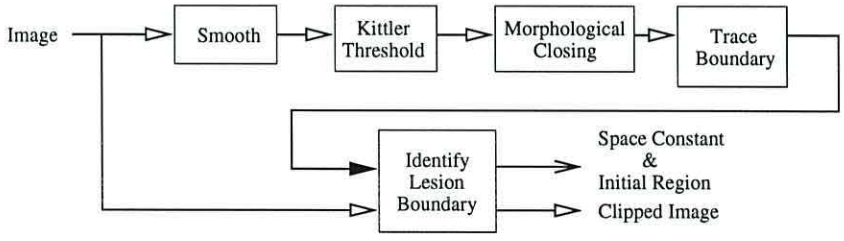


Figure 4.1: Preprocessor: find most lesion-like object.

A later version of this stage first applies tilt removal by least-squares fitting a

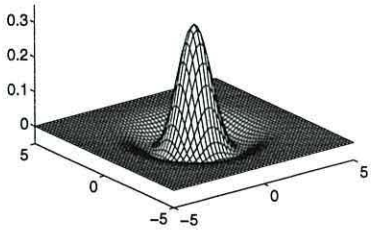
plane to the image, and subtracting a zero-mean version of this from the original image. This helps reduce some problems with the thresholding process caused by inconsistent illumination, however the research in this study did not include this stage.

Edge Sequence Generation

The essence of the edge focusing process is contained in the edge sequence generation stage. The sequence is generated by an iterative process which begins by detecting large scale edges and then reduces the scale of detection step by step, but searching for edges *only* close to that found in the previous step. There are two reasons for limiting the search in this manner: firstly the computational burden of the actual edge detection is reduced, and secondly the confusion of edges that results from the sensitivity of a smaller scale detector to small image 'noise' features is avoided. The result is a sequence of edges that adhere to increasingly fine variations in the lesion outline.

This conceptual process can be used with any scalable edge detector. Two such detectors, LoG and Canny [56] were used with lesion images in [16] and both were found to be effective. The LoG detector was used to create the boundaries used in this study. The LoG edge detector consists of two stages, convolution of the image with a LoG filter followed by identification of zero-crossings. The Laplacian is a second order differential filter and combination with a Gaussian 'window' results in a radially symmetric two-dimensional kernel which finds the edges in a smoothed version of the image. The LoG filter kernel is defined as follows and the two dimensional representation to the right clearly shows its characteristic 'Mexican hat' shape:

$$\mathrm{LoG}(x,y) = \frac{1}{2\pi\sigma_{sc}^4}\left(2 - \frac{x^2+y^2}{\sigma_{sc}}\right) e^{-\left(\frac{x^2+y^2}{2\sigma_{sc}}\right)}$$



The second order nature of the detector means that edges are indicated by locations where the output changes from positive to negative values, called zero-crossings.

Figure 4.2 illustrates the process of edge sequence generation using the LoG edge detector. LoG convolution and zero-crossing detection are denoted by LoG and ZC. The detection is followed by edge tracing to reduce the image information to coordinate form and edge cleaning to provide a single closed boundary, these two steps are respectively denoted by TR and CL. GenM denotes the conversion of boundary coordinates into the two mask image types (fA and fB) required for the next iteration.

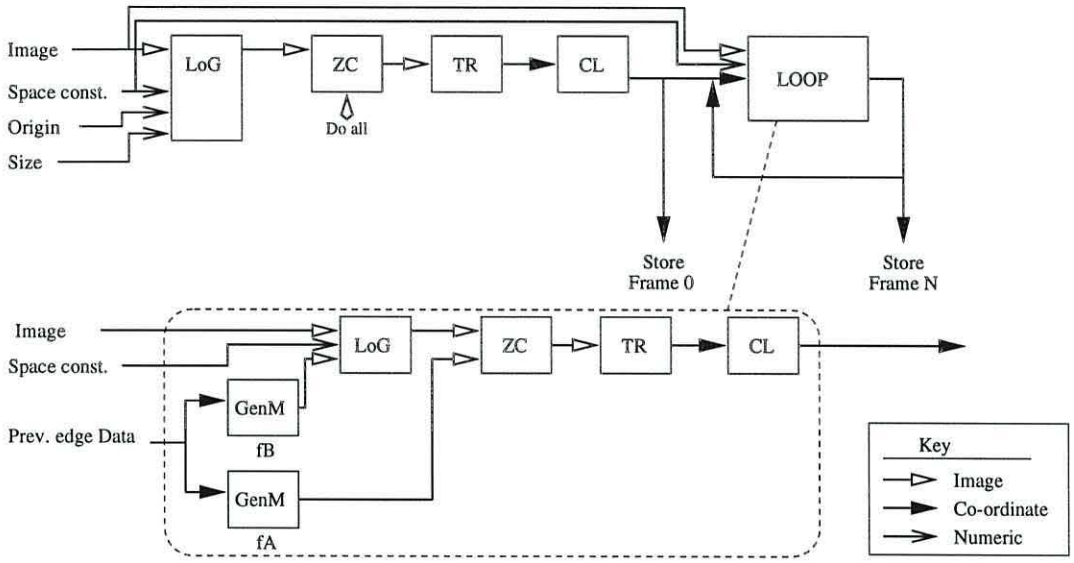


Figure 4.2: Edge estimate sequence generation.

The initial scale of the LoG filter is calculated so that only features of the size as estimated by the preprocessing stage for the lesion will be detected. The process works on a sub-image provided by the preprocessor which is centred on the identified lesion object and which is DC padded (replication of the outermost pixels of the image) so that the initial LoG filter can correctly function for all the pixels near to the preprocessor's estimate of the lesion boundary. Subsequent 'frames' in the edge sequence use filters on a linearly decreasing scale (the space constant σ_{sc} is reduced by 0.5) chosen so that theoretically the edge will move at most one pixel between each frame.

Boundary Selection

The boundary selection stage simply aims to identify the optimal edge in the sequence. The notion of optimality in terms of edges for lesions is not a trivial

matter as discussed earlier. However, the simple definition of 'boundary quality' as the degree of separation between dark lesion and light skin allows the automatic selection of a visually sensible frame from the sequence. This definition is also reasonable considering that the edge focusing process is designed to find the lesion boundary in a grey-scale intensity image based on rapid change in image value (the notion of an 'edge').

Figure 4.3 gives a representation of the process applied to each frame in the sequence. The boundary quality estimate is calculated as the ratio of the mean value for the strip of pixels just inside the boundary and the mean value for the strip of pixels just outside the boundary. The minimum of these values then indicates the optimal boundary.

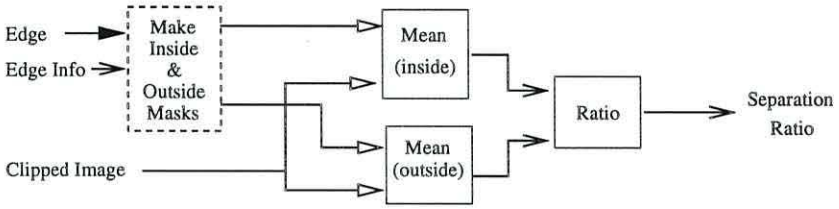


Figure 4.3: Boundary selection: select optimal edge from sequence.

4.3 Porting To Khoros

Image processing systems, regardless of their application domain, often essentially consist of a number of elemental processes, such as edge detection or image smoothing, applied in a certain order to an input image. Moreover, a substantial proportion of these processes can be achieved through various methodologies, for example, edge detection via Laplacian of Gaussian, Sobel, or texture operators and pixel classification by K-means, Bayesian statistical analysis or various neural network architectures . The advantages of developing a system in an environment designed on the principle of data flow between stand-alone processing elements are therefore clear. This principle, together with automatic code management, a library of image processing elements and a visual programming language form the basis of the Khoros [22, 57] environment.

In order to facilitate both the further development, improvement and evaluation, as well as the integration into further work, of the edge focusing technique, the existing system was reconstructed within the Khoros environment[1] using its visual programming environment, Cantata [23]. As a result, the entire process is decomposed and presented in a highly visual manner with the basic processing actions linked by data flow lines. This presentation is ideal for experimentation since it is not only highly modular and more intuitively accessible, but also allows elements of the process to be readily exchanged, allowing for simple evaluation of alternatives and for reuse of elements in future projects.

The Re-Implementation Process

The process of re-implementation within Khoros' Cantata follows the pattern of: selecting *elements*, defining their interfaces, encapsulating them within Khoros *glyphs* and finally linking these together to reconstruct process as a Cantata workspace.

As discussed above, The edge focusing system can be broadly divided into three stages; preprocessing, generation of an edge estimate sequence, and optimal edge selection. These stages are intuitively separate being pre, main and post processing and were implemented as separate workspaces. Within each of these stages the system must be decomposed into units that represent the various elements of the process. These units can then be presented as Khoros *glyphs* from which the process stages can be rebuilt by linking them together to indicate the data-flow.

The process of deciding the extent (functionality content) of each glyph is analogous to designing a flow-chart; too much in one element and understanding is impaired and the workspace will not display its own function, too little and the flow becomes over complex and the overall picture cannot be seen. An additional consideration for glyphs is the reduction in re-use potential for large (and hence highly specific) glyphs versus the overhead incurred in data transfer between overly small glyphs.

The edge focusing stage illustrations in figures 4.1, 4.2 and 4.3 of section 4.2 are in fact directly modeled on the workspaces which resulted from re-implementation

[1] Khoros II (version 2.2.0) was in use at time of writing

of the system within Khoros. Figure 4.4 shows the actual appearance within Cantata of the edge estimate sequence generation stage — the contents of the loop (corresponding to the lower half of the process illustration in figure 4.2) are displayed by clicking on the 'While Loop' glyph.

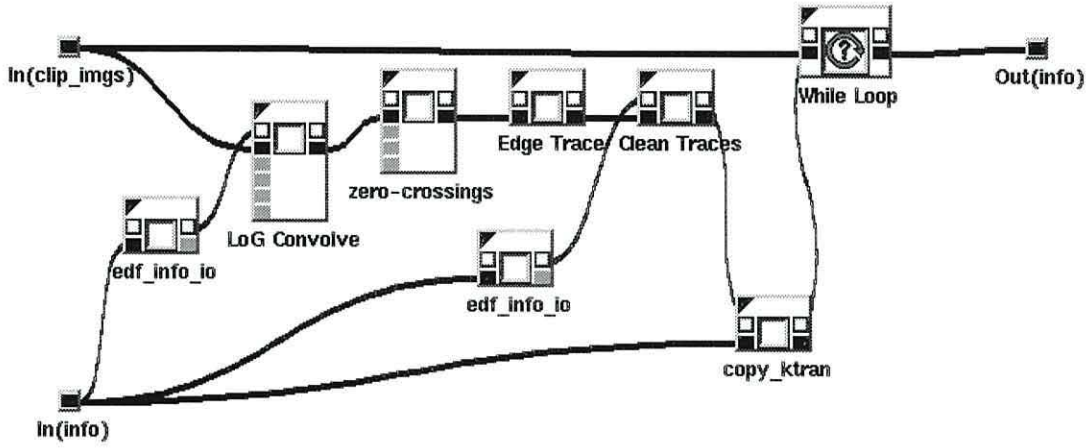


Figure 4.4: The Khoros workspace for edge estimate sequence generation.

Many of the elements required for the conversion of the preprocessor stage exist as pre-defined glyphs from the Khoros standard distribution. New glyphs were required for the Kittler thresholding algorithm and the heuristic lesion object selector, Identify Lesion Boundary. The existing C code for Kittler thresholding was inserted into a glyph simply using the Khoros data transport functions as a shell surrounding it. The lesion object selector however, was re-implemented in order to allow for greater flexibility and transparency in the heuristic process. This decision was proved to be advantageous when, during final testing stages, it was discovered that changing the final selection heuristic (from 'smallest bulkiness' to 'most central' of those regions remaining) yielded a greatly increased robustness in identification of the lesion object.

The edge sequence generation stage required four main elements to perform the LoG convolution, zero-crossing detection, edge tracing and edge cleaning. These were created by wrapping the relevant C code from the existing system with Khoros data transport functions as with the Kittler thresholding. The only other element or glyph required was one which could generate the required masks for the

next iteration from the coordinate file output of CL. Such a tool had already been partially devised for the testing of the glyphs mentioned above. The addition of some control over the draw format was all that was required to make the existing glyph perform the required task.

All the elements required for the optimal edge selection stage were available as pre-defined Khoros glyphs or had been created for use in the other stages.

4.4 Discussion

This section begins by evaluating the Khoros implementation of the edge focusing system both in terms of results and performance relative to the existing system, as well as in relation to the goals motivating the conversion. Next the performance in relation to the image set used in this study of the edge focusing technique *in general* is considered and particular limitations are highlighted. Some possibilities for improvements to this general performance are then discussed, and finally the boundary definition policy used in the remainder of this study is presented.

4.4.1 Evaluation of the Khoros Conversion

Evaluation of the reconstructed system was undertaken using two differing measures. Firstly results and stability were compared with the existing system and secondly the envisaged improvements in adaptability, accessibility to experimental change, and reuse potential were investigated.

Over a range of initial input images, the new system was seen to consistently duplicate the results obtained from the original both in terms of the generated edge sequence and the selection of best boundary made. However, the new system was seen to be slightly less stable than the original – failures occurred in some cases resulting from underestimation of lesion object size by the pre-processing section. A variety of contributing factors to this underestimation were identified, an example being a different definitions of morphological operators in the Khoros

contributed toolbox MMACH.

The new system was appreciably slower than the original, mostly due to the Khoros data-flow paradigm resulting in the reading and writing to files between each glyph element. The implementation of the edge sequence generation stage has however been improved over that discussed in [23] which 'stacked' the coordinate information for each frame into a single file. The reading and writing of this growing file for each frame was found to involve a considerable time penalty, the current implementation (termed non-stacking, ns) avoids this by using separate files whose name contains the frame number. The implementation of the optimal selection stage did not require the creation of any new glyphs, however it does consequently involve a large number of glyphs so that the execution time for the stage is dominated by file transport between glyphs. The speed of this stage could be dramatically improved by merging the functionality of a number of these glyphs into a new single glyph. The general speed issue has also been addressed directly in the current version of Khoros which now allows inter-glyph data flow to be carried out using faster transports such as pipes and shared memory, use of which may well considerably reduce this speed difference.

In terms of the envisaged accessibility improvements, the process of testing repeatedly demonstrated the advantages of access to all data passed between the glyphs — the faster, but transitory, transports now available in Khoros are not used in the current version for this reason. The availability of a variety of data visualization and analysis tools which can simply be 'plugged in' to the output of any glyph greatly increases development prototyping and testing turnover.

The displayed workspaces describe the method for the system they contain nearly as well as a separately prepared flow-chart, which is a great advantage in terms of reuse and development potential for the system. In addition, a number of the glyphs developed during this project have already been reused in other research undertaken by this group.

In all the new system satisfies the aim of making the edge focusing boundary generation system much more accessible for use in the development of the envisaged lesion diagnosis support package.

4.4.2 Limitations of the Edge Focusing Technique

The difficulties involved in providing even a reliable definition of what constitutes the boundary of a lesion make generalized evaluations of the performance of the edge focusing technique difficult. The generation of synthetic lesion images, where the true border location is a known parameter of the generation process, therefore provides an important tool for detailed evaluation of performance [18] and this technique has been used to supplement trials on real lesion images in the research which originally produced the lesion boundary edge focusing method [16]. This combination of evaluating both synthetic images in relation to their known boundaries, and real images in relation to manually defined boundaries is perhaps the most reliable method available.

The discussion here however, concentrates on the performance of the edge focusing technique in relation to the particular image set used in this study. Four main problems exist for the technique:

Not enough surrounding skin. Some of the images used in this work contain lesions whose borders come very close to, or even touch the edge of the image. As noted in the description of the edge focusing algorithm earlier in this chapter, the proper function of the LoG filter requires a sufficient area of data surrounding the boundary location so that the whole filter, when centred on the boundary location, will fit into the image. In most cases where the lesion only comes *close* to the edge of the image, the use of DC padding allows the process to proceed effectively, however in cases where the boundary is too close to the image edge the gradient information is critically corrupted by the padding and where the boundary actually touches the edge of the image padding is always ineffective. The malignant melanoma shown in figure 4.5 top left is such a case.

Lesion boundary poorly defined. Lesion boundaries are not always well defined 'edges'. The proper function of any edge detection based boundary finding method requires that the boundary can be seen as a clear gradient in the image data being analysed. Furthermore, correct localization of the border requires the consistent identification of a feature of the edge (for

example its maximum gradient) with the lesion boundary. Lesions such as the benign naevus shown in figure 4.5 (top right) with 'fade in' boundaries can cause difficulties relating to these requirements: The smoother hand drawn boundary (shown in green) is the outline of the discernibly different light brown lesion area, which is considerably different in some areas to the generated boundary (shown in red).

Lesion 'colours'. Lesions often show a number of different colours, for example, variegated colouring is a widely recognized indicator of melanoma, halo naevi derive their name from their de-pigmented border and lesions may well show inflamed borders. Different colours cause problems both because edges exist between each area of differing colour, and in normal grey-scale intensity images some different colours are seen as the same grey level. The benign naevus shown in figure 4.5 centre left and centre right, provides a good illustration of such problems. In the colour version, the edge of the central dark brown area is more pronounced than the edge of the entire lesion. Furthermore whilst the pink area to the left of centre is clearly distinguishable from the light brown below it in the colour version, in the grey-scale version there is no apparent difference. In addition some lesions clearly extend beyond the pigmented area, as in figure 4.5 bottom left.

Non-uniform illumination and loss of focus. Non-uniform illumination can cause problems in boundary detection as the pixels corresponding to a single image area such as the skin can, as a result, show 'false' variation in intensity. The top left image in figure 4.5 has a clearly brighter band across the centre. The original lesion image edge focus process used illumination tilt compensation to tackle this problem, however it is evident that the technique as used, which fitted a plane to the illumination, would not be successful in cases similar to the example which would require the fitting of a more flexible surface. The band of increased brightness in the example could have been the product of curvature in the subject, such as resulting from the lesion being located on the arm or leg. Subject curvature can also cause loss of focus, and consequently loss of fine detail in areas of the image. Although this is generally not problematic for techniques such as the LoG edge detector which employ smoothing and therefore remove the fine detail anyway, it can cause problems for pixel similarity segmentation

techniques as the different areas will fade into each-other, for small-scale texture techniques as the properties of pixel value variations are changed by blurring, and for large-scale texture techniques such as that presented later, which analyses fine linear features which are lost by blurring.

Image 'noise' features – hairs. Image features also exist such as scales, rulers and hairs, which essentially can be classified as noise in terms of boundary detection. For the image set used here the only such feature that has proved significant has been the presence of hairs. Dark hairs are seen as highly significant edges and consequently they cause 'glitches' in the line of the boundary where the detector attempts to include the linear hair feature as part of the lesion.

4.4.3 Improvements

The first of the identified problems can only be realistically solved by better control in the acquisition of the images. The resolution of even relatively inexpensive modern digital cameras would allow the use of a relatively wide field of view whilst maintaining the level of spatial resolution used in this research. The full edge focusing technique is designed to be tolerant to image features such as limb edges which may result from the adoption of a wider field of view. The addition of a means of identifying the skin areas would of course be necessary for skin/lesion differential measures, however simple procedural control at the capture stage (such as the use of a bright blue cloth backdrop) should make this information readily accessible.

There is no simple solution to the lack of a good definition for what constitutes the edge of a lesion.

The problem of different lesion colours can be addressed by the development of a better mapping from colour information to values which more clearly represent the lesion/skin distinction. Several such transforms have been proposed as mentioned earlier and discussed in detail in the Colour chapter. However, it is not clear whether a 'perfect' transform even exists and furthermore, the problem of part-

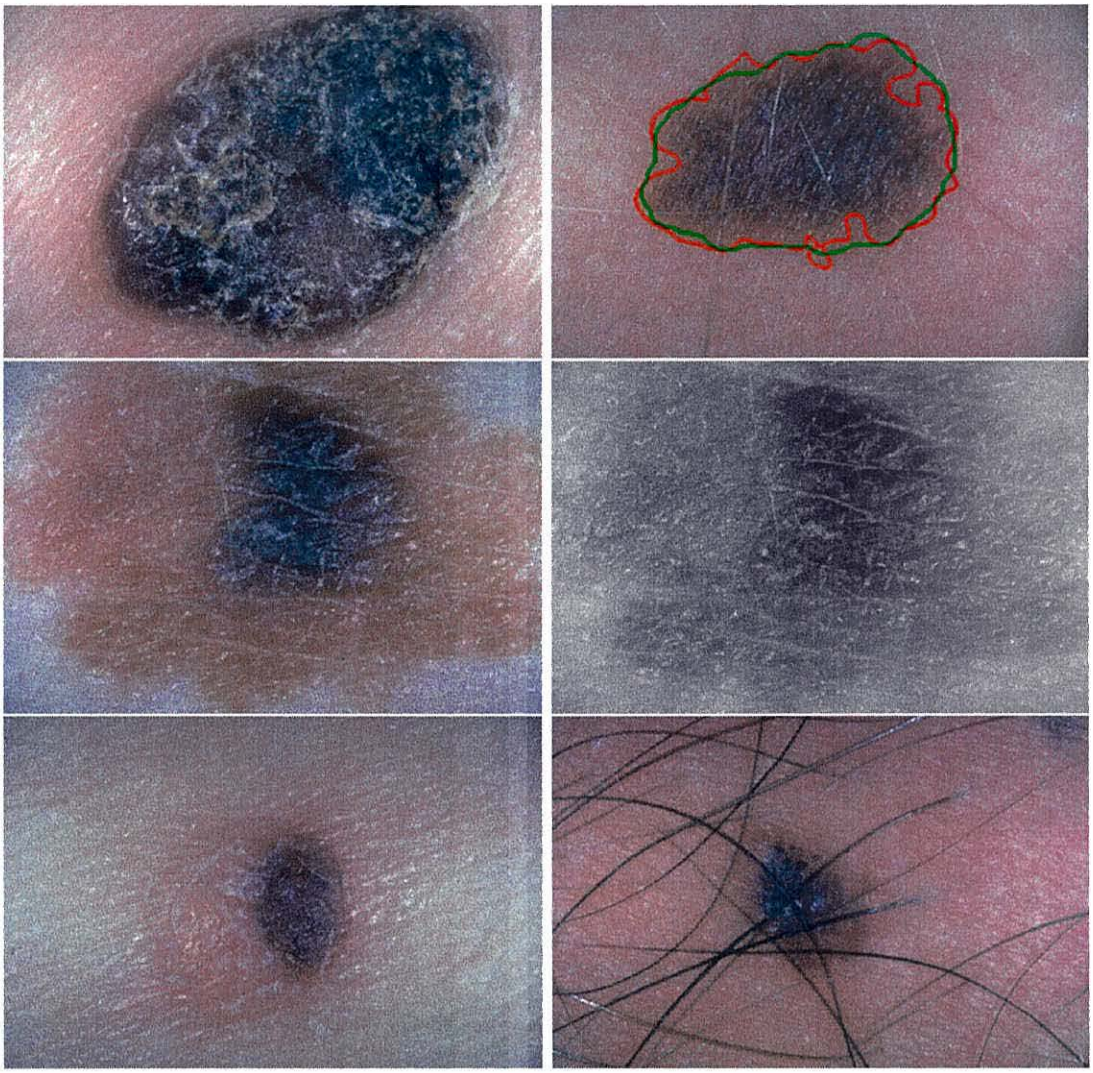


Figure 4.5: Example images illustrating performance issues for the edge focusing technique.

pigmented lesions still remains. It is likely that a multi-channel approach merging boundaries obtained from different initial data (edge, colour, texture and possibly even other image modalities such as UV, IR and ultrasound) would be required to obtain a truly reliable boundary.

Non-uniform illumination could be addressed through illumination surface correction. As identified above, a simple planar surface as used in the original edge focus process would not always be sufficient, however more flexible surfaces could cause problems by 'correcting' for the darker lesion area as an illumination surface artifact. Loss of focus may be correctible using techniques such as de-convolution,

however this would involve calculation of the correct point-spread function as it varies over an image. Neither of these issues could be investigated in any detail in the time available for this study.

The difficulties which can be caused by hairs obscuring lesion images has been recognized by Lee *et al* [58] who propose the *Dull Razor* algorithm as a possible solution to this problem. In essence the process simply locates strong and extended linear features and then uses interpolation between values either side of the hair to 'erase' the hair from the image. The linear feature detection is performed using grey-scale morphological closing with linear structuring elements oriented at 0°, 45° and 90° combined by taking the maximum of these three results. The red, green and blue bands are processed separately and the 'hair mask' for band x, M_x is defined by thresholding the absolute difference between the combined morphological result and the original values for the band. The final mask M is the union of M_R, M_G, and M_B. M is then refined by accepting only extended linear regions. The length of the longest line segment within M, for each of the eight major directions radiating from each pixel in M is calculated. Only pixels for which the maximum of these lengths is greater than 50 and the minimum less than 10 pixels are accepted. The hair pixels defined by M are then replaced by interpolation between the value 11 pixels beyond either side of the hair area along the direction of the shortest length. In the resulting image at this stage thin and relatively faint lines are still often seen marking each side of the hairs. These are removed by performing adaptive median filtering for all pixels within M dilated by a 5×5 square structuring element.

Figure 4.6 shows an example of Dull Razor application to a grey-scale image. The original image (left) has a number of significant hairs which are removed effectively in the processed version (right). The use of Dull Razor, in combination with the research into colour analysis, was not pursued as the impact of hairs in the images used in that work was not significant. The smoothing performed as part of the Dull Razor process made this technique inapplicable to the work on texture analysis as it removed the fine detail needed for that analysis in some parts of the image. This can clearly be seen in the processed example shown, especially in the skin to the left of the lesion. In terms of future work however, the impact of Dull Razor as a pre-processing step for the edge focusing technique

should nevertheless be addressed.

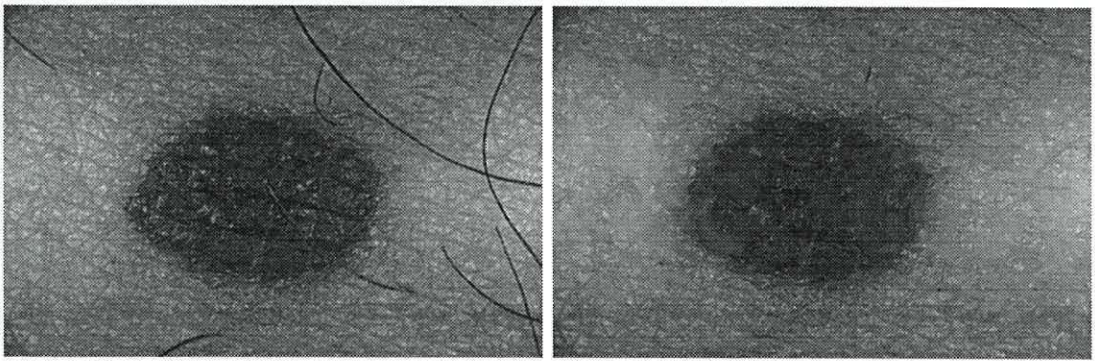


Figure 4.6: An example of Dull Razor hair removal processing.

4.4.4 Boundary Definition Policy

Boundary information is used in both the colour and texture analysis sections of this study. In the former the boundaries are used to identify lesion and skin in the analysis of colour properties and in the latter for lesion/skin comparative evaluation of a texture based feature.

The boundaries used in this study were, wherever possible, generated using the edge focusing system (using the Khoros version and starting with standard grey-scale intensity image). However, the problems outlined above meant that some boundaries were not obtainable via this automatic method. The automatic process was deemed to have 'failed' where the edge focusing process could not be applied (such as where the lesion touches the edge of the image) or the final boundary was considerably different from that of a visual estimate as defined below (as can result from, for example, indistinct boundaries and boundaries which come too close to the edge of the image). In failure cases a boundary has been constructed by hand from grey-scale intensity images so as to correspond with the the limit of the first reasonably different intensity from that of the surrounding skin. This means that the lesion is identified as the extent of the noticeably darker grey area (including any areas of regression inside the notional lesion). Grey-scale intensity images were used so as to better resemble the results obtained using the current implementation of the edge focusing system.

The construction of boundaries by hand is both highly time consuming and naturally prone to problems of subjectivity and inconsistency, with the final result often less faithful to small scale boundary irregularities. The part automatic, part hand constructed boundary set is sufficient for the purposes of this work as the extreme accuracy and consistency required to perform analysis such as boundary irregularity are not vital for the colour property analysis, and small variations in boundary location are likely only to produce small variations in the comparative texture feature analysis presented in this study.

4.5 Conclusions

Robust and accurate identification of the extent of a lesion is a vital element of any image based skin lesion diagnosis support system. Boundary information directly enables the estimation of a variety of important diagnostic indicators such as size, shape, asymmetry and irregularity. In addition, and directly relating to the main body of the research presented in this study, boundary information enables the evaluation of features and properties both for the lesion area in isolation, and comparatively between the lesion and surrounding skin.

Boundaries drawn by hand suffer from both inter and intra-observer inconsistency and are further hampered by the difficulty in providing even a good definition of exactly what constitutes the boundary of a lesion. Naturally the desire for an automated boundary detection method is strong and consequently much research has been directed toward this aim. In previous research, the edge focusing technique has been proposed and shown to be effective and reasonably robust as a means of accurate automatic lesion boundary detection.

The existing edge focusing system was reconstructed in the Khoros II environment in order to facilitate the further development, improvement, evaluation and integration of the technique. To achieve this the entire process was decomposed into its functional elements and then rebuilt within the Cantata visual programming environment of Khoros. Within Khoros, the implementation the edge focusing process is presented in a highly visual manner with the basic processing actions

linked by data flow lines. Evaluation in terms of the goals driving the conversion have shown the value of this presentation; it is ideal for experimentation since elements can be readily exchanged, allowing for simple evaluation of alternatives and for reuse of elements in future projects, and it greatly increases development prototyping and testing turnover by allowing access to all data passed between the process elements and by providing a variety of data visualization and analysis tools which can simply be 'plugged in' to examine the data at any point. The re-implemented system has a significantly longer execution time due to the overhead involved in data flow using permanent files. However if speed becomes an important issue, Khoros provides the option of using much faster, non-permanent (but consequently non-interrogable) data transports.

In general the edge focusing technique is effective in providing robust and accurate lesion boundaries, however certain situations (such as lesions with highly indistinct boundaries and lesions which extend to the edge of the image) can cause poor results or failure. Possible solutions are proposed for some of the identified problems including the use of the Dull Razor technique for removing hair features from the image. The boundaries used in this study are governed by a boundary definition policy which is in essence that: boundaries are obtained, wherever possible, from the edge focusing system, but where the automatic system is ineffective or fails, boundaries are constructed by hand.

Chapter 5

Colour

This chapter details the research undertaken into the possibilities for obtaining diagnostic feature information through the use of colour data. The importance of colour in diagnosis of skin lesions is widely recognized and reflected in the inclusion of colour features in both skin cancer checklists and in differential diagnosis descriptions. The concept of colour is introduced and the implications in terms of colour model resulting from requirements of the envisaged diagnosis-support system are discussed. The particular features pursued in image based diagnosis of skin lesions are identified together with the different methodologies used in their detection and computer based analysis. Current research into colour image analysis (the methodology most suited to the envisaged system) particularly for skin lesion images and their segmentation is reviewed. The underlying goals of segmentation, particularly in relation to lesion images, are investigated in detail and conclusions are drawn resulting in the development of a new region-based technique. Initial results prompt a detailed consideration of other colour-spaces together with suitable colour similarity measures and patterns of lesion data distribution for them. Finally the segmentation performance on the new transformed data is presented and discussed.

5.1 Colour Imaging

Although colour seems at first to be a simple concept, expression of exactly what the term means in practice is far from simple. In an abstracted sense the colour of an object is the combination of the intensity of reflected radiation over the range of frequencies known as visible light. The colour of an object is therefore a product of both the intensity of illumination across this entire spectrum coupled with the reflectance and scattering properties of the materials from which the object is composed.

Consideration of colour in terms of such a model of complex interactions would however be awkward and confusing for the purpose of the discussions which follow. A much simpler model will be used in which colour is described as a combination of intensity of reflected light in the three bands of red, green and blue (RGB). This model reflects human colour perception and is widely used in the representation of colour images and especially so for digital imaging and computer based applications.

The requirement, in the overall diagnosis support programme, for a low cost system indicates the need to use inexpensive commercial (digital camera) technology in image acquisition. Such digital cameras are obviously designed and developed with the aim of 'accurate' reproduction of a scene where accuracy simply relates to human perception. The starting point for this investigation of colour analysis for lesion diagnosis is consequently assumed to be a digital RGB image whose 'quality' relates only to visual fidelity in reproduction of the scene. In other words it will be assumed that the images 'look like' the lesion being imaged.

5.2 Colour-spaces

A colour-space is a model which is used to specify colours in a useful and standardized manner. Many colour-spaces have been used in image processing and the most useful representation of colour is often directly task dependent; the use of the RGB representation is clearly valuable where hardware is concerned,

whereas other spaces such as IHS and $L^*u^*v^*$ (described in detail below) more closely relate to a human characterization of colour properties.

Decoupling

As previously described, the RGB colour-space defines colours by the intensity of light in the three bands of red, green and blue. It is often much more useful to decouple the "brightness" of a colour from the underlying "colour". The brightness is often described by the terms *intensity* or *luminosity* and the underlying colour as *chromaticity*.

Intensity - Chromaticity

The intensity is defined as the mean of the red, green and blue values. This relates to the overall stimulation of the three colour detectors of the human eye. The chromaticity can then be described by the relative levels of each of the red, green and blue intensities, which are then written as lower case r, g and b. Using the definitions (given in equations 5.1, 5.2, 5.3 and 5.4) it is clear that $r+g+b=1$ so that these three coordinates describe a plane in their 3D space and that only two of the three is necessary for complete description (e.g. $b=1-(r+g)$). As a consequence chromaticity can also be described with 2D coordinates (which will be called cX and cY or together as *chromaXY* here) as given in equations 5.5 and 5.6.

$$I = \frac{1}{3}(R+G+B) \tag{5.1}$$

$$r = \frac{R}{(R+G+B)} \tag{5.2}$$

$$g = \frac{G}{(R+G+B)} \tag{5.3}$$

$$b = \frac{B}{(R+G+B)} \tag{5.4}$$

$$cX = r - \frac{1}{2}(g+b) \tag{5.5}$$

$$cY = \frac{\sqrt{3}}{2}(g-b) \tag{5.6}$$

note: this chromaXY definition has been chosen to orient the data to match that produced by the IHS conversion given in this document.

The chromaticity coordinates describe a triangular space with the primary colours at the three vertices and the secondary colours at the mid-points of each side. The pure colour wavelengths are represented at the edge of the triangle and moving towards the centre that same colour is increasingly diluted with achromatic (white) colour. Figure 5.1 shows (left) a conceptual view of the intensity-chromaticity space with its bi-tetrahedral shape and (right) an illustration of the chromaXY space formed by taking colour values spaced equally throughout the RGB cube and mapping them to the chromaXY space.

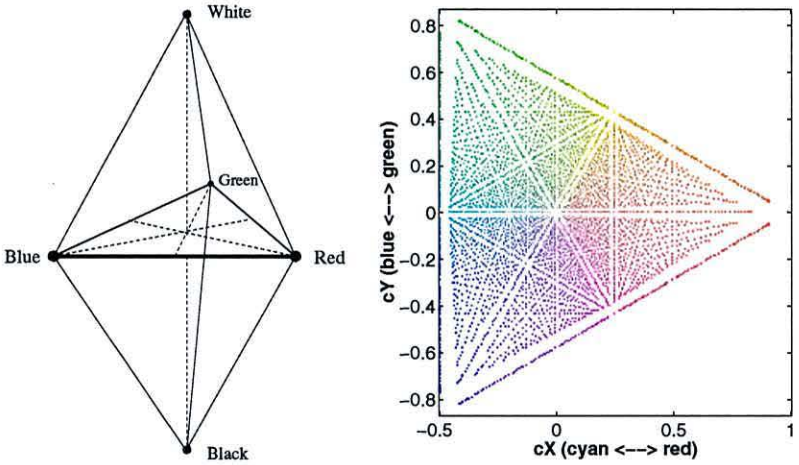


Figure 5.1: The intensity-chromaticity space: Left, a conceptual view of the whole space, and right, a colour view of the chromaticity space

IHS Intensity Hue and Saturation

The IHS colour-space is an obvious development from the intensity-chromaticity model. The I stands for Intensity and H and S for Hue and Saturation respectively. Hue and saturation provide a clearer description of chromaticity where the dominant colour is decoupled from the "purity" of the colour. Fully saturated colours are those of single wavelengths and zero saturation indicates achromatic colour. Considering the chromaticity triangle, if w is the white point at the centre, p the given colour, and p_s the fully saturated version of that colour (the point where a line from w through p intersects the edge of the triangle), then the saturation is given by the ratio of the distance $w \rightarrow p$ and $w \rightarrow p_s$. The Hue represents the dominant wavelength as an angle which by convention has red at 0° and moves through green at 120° and blue at 240° and then through purples back to red. IHS values can be calculated as given in equations 5.7, 5.8 and 5.9

which provide values in the same range as the input data for I, $0 - 2\pi$ radians for H, and $0 - 1$ for S.

$$I = \frac{1}{3}(R+G+B) \tag{5.7}$$

$$H = \begin{cases} H' & \text{where } B/I \leq G/I \\ 2\pi - H' & \text{otherwise} \end{cases} \tag{5.8}$$

$$\text{where } H' = \cos^{-1}\left[\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}\right]$$

$$S = 1 - \frac{\min\{R,G,B\}}{I} \tag{5.9}$$

Technically H remains undefined if $S = 0$ and S remains undefined if $I = 0$. However, in order that all values may be defined for the purposes of the processing in this work, $H = 0$ and $S = 0$ respectively are substituted in these cases.

[59]

The derivation of the formulae for IHS is not trivial. Many sources use $\tan^{-1}$ on two dimensional x,y chromaticity coordinates (which can be simply derived from r,g,b chromaticity coordinates) in the calculation of H, ignoring the fact that this form requires many special cases. The S value is also often wrongly derived, for example, in Pratt [60] the saturation is given as the simple distance between the white point ((0,0) in x,y chromaticity coordinates) and the (x,y) point for the given colour, rather than the ratio of this length to that of the equivalent saturated colour – the practical result of which being that a fully saturated secondary colour such as yellow would be attributed with a S figure of 1/2. The formulae given here are presented with full derivation in Gonzalez and Woods [59].

Figure 5.2 (left) shows a conceptual view of the IHS space showing the cylinder form. A comparison of this diagram and that for intensity-chromaticity in figure 5.1 shows that this space will be increasingly less densely populated moving toward either extreme of I, with zero and maximum intensity both having only one possible point representing black and white respectively. Figure 5.2 (right) gives an illustration of the IHS Hue-Saturation space again formed by taking colour values spaced equally throughout the RGB cube and mapping them to IHS.

It is important to recognise that the use of an angle to represent Hue requires a colour similarity measure on IHS space to account for the effective identity of

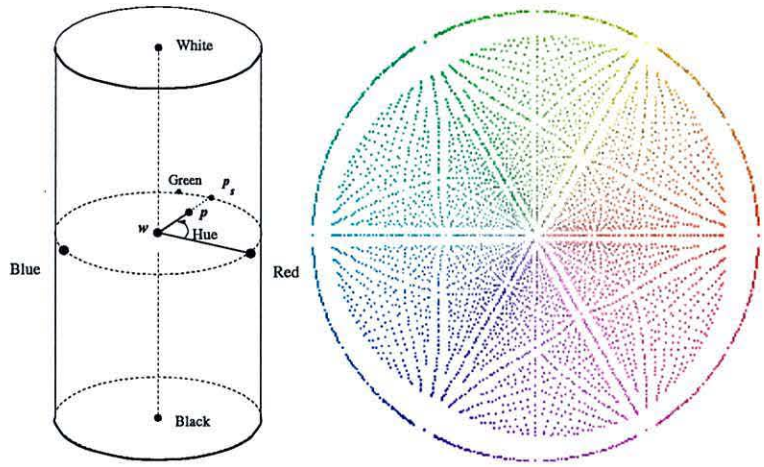


Figure 5.2: The IHS space: Left, a conceptual view of the whole space, and right, a colour view of the Hue-Saturation space

0° and 360°. In addition it is unclear exactly what distance measure should be used to reflect the properties of the human visual system for this space, or even if any simple measure could be constructed, however the IHS colour-space is a widely used and powerful representation of colour data [59] so that the properties of lesion image colour information in this space should be investigated.

CIE $L^*u^*v^*$

The $L^*u^*v^*$ colour-space [60, 61] was developed by the Commission Internationale d'Eclairage (CIE) and became a CIE standard in 1976. $L^*u^*v^*$ evolved from the $L^*a^*b^*$ and $U^*V^*W^*$ colour-spaces (definitions for these can be found in [60]) specifically with the aim that the distance between two colours would directly reflect human perception of colour difference.

The construction of such a uniform colour-space involves the concept of a 'just noticeable colour difference' (*jncd*). Plotting the locus of 1 *jncd* about a certain colour generally forms an ellipsoid (often referred to as a MacAdam's ellipse). In a uniform colour-space, all such ellipsoids should be spheres. It is known that, in general, no linear transform to such a uniform space exists [62] (although a complex non-linear transform has been developed [60]).

$L^*u^*v^*$ is the CIE standard for a uniform colour-space where the colour difference

is defined as in equation 5.10

$$\Delta^2\{C_1, C_2\} = (\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2 \tag{5.10}$$

Which is simply the L2 distance between two colours in that space.

The conversion from RGB to $L^*u^*v^*$ first requires transformation of the RGB values to CIE XYZ. Both RGB and CIE XYZ are three-primary or *tristimulus* colour systems. RGB uses the three real primaries, red, green and blue (the exact definition of these primaries determines the particular "flavour" of RGB). Positive combinations of real primaries are however only able to produce a subset (the gamut of the primaries) of all visible colours – those contained within the triangle formed by these primaries – and this situation is exacerbated by the need for these to be available as phosphor colours. The XYZ system uses three artificial primaries $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ chosen such that all visible colours fall within their gamut and the Y value is equivalent to the luminance of the colour. Figure 5.3 shows the gamut of RGB_{NTSC}[1] and XYZ colours together with that of all visible colours. The curving line is the locus of real pure colour wavelengths (red≈700nm, green≈546nm and blue≈436nm) and encloses the gamut of real colours.

Conversion between tristimulus systems can always be achieved by a simple matrix operation. Equation 5.11 gives the transformation for RGB_{709D65}[2]. RGB_{709D65} is used here as the lesion images used in this work are stored on Kodak Photo-CD, the digitization process for which uses ITU Rec 709 and D65 reference white [64].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412411 & 0.357585 & 0.180454 \\ 0.212649 & 0.715169 & 0.072182 \\ 0.019332 & 0.119195 & 0.950390 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5.11}$$

This assumes RGB_{709D65}

[63]

The derivation of this transformation matrix is as follows:

[1] National Television Systems Committee (NTSC) receiver phosphor standard

[2] RGB_{709D65} is used here to stand for International Telecommunication Union (ITU) (http://www.itu.ch/) Recommendation 709 primaries with D65 reference white. ITU Rec 709 is the new designation for CCIR 709, after the CCIR (Comite Consultatif International des Radiocommunications) was absorbed into its parent body the ITU [63]

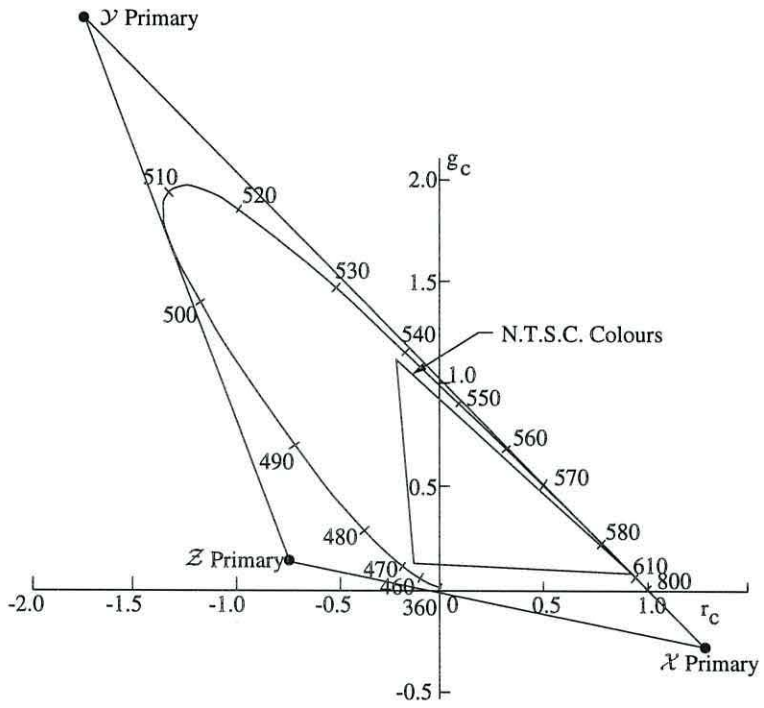


Figure 5.3: Chromaticity diagram showing the gamut for XYZ RGB_{NTSC} and real colours. (NOTE: in this diagram the axes represent red and green chromaticity coordinates in terms of CIE 1931 standard primaries. The NTSC gamut includes negative values as that system uses the different NTSC primaries.) (adapted from Pratt [60])

The CIE xyY space is derived from the XYZ space and the xyz chromaticity coordinates of this space. Chromaticity coordinates are the normalized values of the tristimulus, i.e. $[x\ y\ z] = \frac{1}{X+Y+Z}[X\ Y\ Z]$.

Since, by definition, $x + y + z = 1$, $z = 1 - (x + y)$, only x and y are required to describe the chromaticity and so, together with the luminosity Y they provide a full description of colour. The XYZ tristimulus values can be recovered from xyY coordinates: $[X\ Y\ Z] = \left[\frac{xY}{y}\ Y\ \frac{zY}{y}\right] = \left[\frac{xY}{y}\ Y\ \frac{(1-(x+y))Y}{y}\right]$.

RGB_{709D65} assumes the D65 (W_{D65}) as the reference white which is defined in xyY coordinates as:

$W_{D65} = [x_0\ y_0\ Y_0] = [3.12713\ 0.329016\ 1]$ (Y is always 1 for white).

The formula above allows the calculation of XYZ values for W_{D65} and since the R=G=B=1 for this white, the transformation matrix

must satisfy the matrix equation:

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} = \begin{bmatrix} x_0/y_0 \\ 1 \\ z_0/y_0 \end{bmatrix} = [m_{i,j}] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{5.12}$$

The $[x\ y\ z]$ chromaticity coordinates of the three primary colours are also needed in the construction of the conversion matrix. RGB_{709D65} uses the ITU Rec 709 primaries:

$$\begin{array}{lcll} \text{red} & = & [x_r\ y_r\ z_r] & = [0.64\ 0.33\ 0.03] \\ \text{green} & = & [x_g\ y_g\ z_g] & = [0.30\ 0.60\ 0.10] \\ \text{blue} & = & [x_b\ y_b\ z_b] & = [0.15\ 0.06\ 0.79] \end{array}$$

By invoking the white balance condition, equation 5.12 can be re-written as:

$$\begin{bmatrix} x_0/y_0 \\ 1 \\ z_0/y_0 \end{bmatrix} = \begin{bmatrix} a_r x_r & a_g x_g & a_b x_b \\ a_r y_r & a_g y_g & a_b y_b \\ a_r z_r & a_g z_g & a_b z_b \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} x_r & x_g & x_b \\ y_r & y_g & y_b \\ z_r & z_g & z_b \end{bmatrix} \begin{bmatrix} a_r \\ a_g \\ a_b \end{bmatrix} \tag{5.13}$$

by substitution of the x's for red, green and blue, a_r a_g and a_b can be calculated and hence the matrix $[m_{i,j}]$ can be found.

[60, 63]

The conversion to the $L^*u^*v^*$ colour-space from XYZ tristimulus values is given by equations 5.14-5.16. The L^* value represents the luminosity of the colour and the u^* and v^* are chromaticity coordinates which for increasing values approximately relate to cyan $\rightarrow$ red and blue $\rightarrow$ green respectively.

$$L^* = \begin{cases} 25\left[100\frac{Y}{Y_0}\right] - 16 & \text{if } \frac{Y}{Y_0} \geq 0.008856 \\ 0.903\frac{Y}{Y_0} & \text{otherwise} \end{cases} \tag{5.14}$$

$$u^* = 13L^*(u' - u'_0) \tag{5.15}$$

$$v^* = 13L^*(v' - v'_0) \tag{5.16}$$

Where:

$$u' = \frac{4X}{X + 15Y + 3Z}$$
$$v' = \frac{9Y}{X + 15Y + 3Z}$$

and u'_0 and v'_0 are obtained by the substitution of the tristimulus values X_0, Y_0 and Z_0 of the reference white.

[60]

Figure 5.4 shows an illustration of the $L^*u^*v^*$ colour-space again formed by taking colour values spaced equally throughout the RGB cube and mapping them to $L^*u^*v^*$. Left is a view of the u^*v^* chromaticity plane which clearly shows the distortions applied to obtain the colour difference uniformity when compared to the other chromaticity diagrams. Centre and right are views of u^* vs L^* and v^* vs L^* respectively. These views clearly show a much smaller range of luminosity when compared with the chromaticity information. Since the L2 distance corresponds (approximately) with human perception of colour difference these relative ranges indicate that the chromaticity information of a colour has a greater importance than the luminosity. This observation has clear implications in terms of the previous colour work in lesion image analysis since, as discussed previously, the dimensionality reduction often applied as the initial step in processing the RGB colour extracts data corresponding approximately to intensity only. Such processing is therefore in effect performing the opposite of the $L^*u^*v^*$ transform in terms of manipulation of the colour difference properties for the output space, by placing increased rather than decreased relative importance of luminosity.

Whilst many other colour-spaces have been developed, the three presented here represent a cross section of the most valuable and widely used spaces. The RGB space is arguably the most widely used colour-space for digital computing, the IHS space has proved immensely valuable in image processing through its separation of 'colour' into its abstract components, and the $L^*u^*v^*$ space is the current international (CIE) standard for a uniform colour-space in relation to human perception.

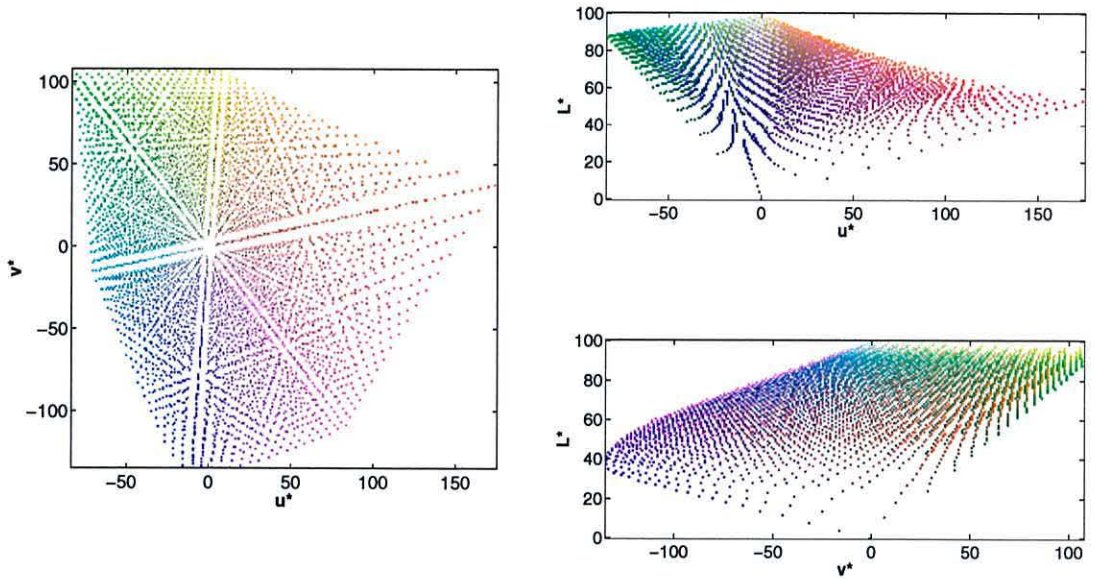


Figure 5.4: The $L^*u^*v^*$ space: Left, the u^*v^* chromaticity plane, right top a view of u^* vs L^* and right bottom a view of v^* vs L^*. The graphs are all to approximately the same scale.

5.3 Colour as a Diagnostic Indicator

Irregularity of pigmentation in a lesion is widely noted as a major feature for the differential diagnosis of melanoma [5, 7] as can clearly be seen by its importance in each of the published checklists described previously. Colour information has consequently been important in research into automated diagnosis. For example, simple methods for quantifying variegated colouring based on variances of each of the three components of RGB together with, amongst other indicators, the colour difference (relative RGB component values) between the lesion and skin areas have been used in trials of automated decision techniques for classification [35, 36]. In addition to irregularity in colouring, the presence of certain colours such as red, white and blue can also provide diagnostic information. The consideration of colour presence seems particularly important in diagnosis based on ELM images where a 'blue-grey (or blue-white) veil' for example is often cited as a significant indicator for melanoma [65–67]. Figure 5.5 contains two melanoma image examples showing variegated colouring in a clinical image (left) and blue-grey

veil in an ELM image (right).

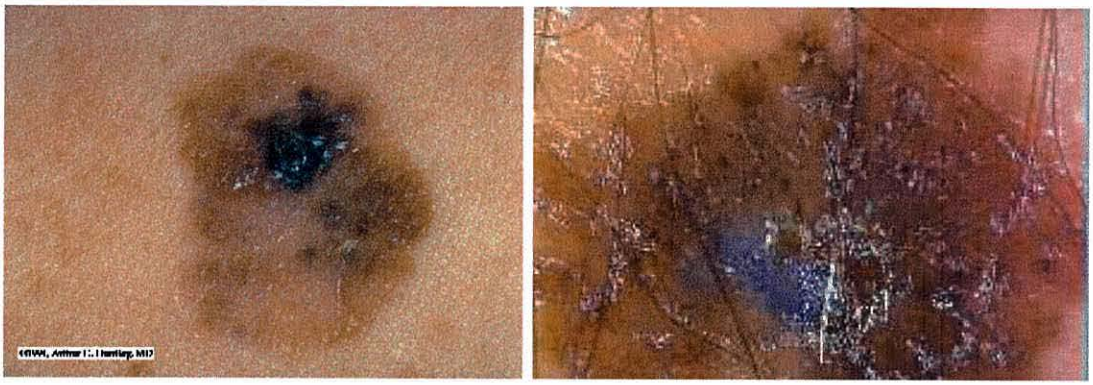


Figure 5.5: Colour in melanoma diagnosis: Left, a clinical image showing colour variegation and right, an ELM image showing blue-grey veil. (from Matrix Dermatology Resources Online [26] and CMIS-CSIRO Online [67] respectively)

Colour can also be used in skin lesion diagnosis through the consideration of the spectral properties of the reflected light. Research of this type can be divided into two categories: In the first, true optical reflectance spectra are obtained for a small number of points for each lesion with the aim of finding common characteristics for a given lesion type. A probe is used to direct light from a known source to a point on the skin surface and to collect the reflected light for analysis. Research in this area has shown that consideration of such spectra can allow discrimination between malignant melanoma and benign naevi [30, 68, 69]. The second category uses the information contained in normal colour images. Analysis of different lesion types is again made, this time with respect to the colour properties (such as the individual or relative intensities of the RGB components) of the image. Consideration of colour images in relation to a light scattering model of the skin has been shown to provide a means of quantification for features such as the level of epidermal melanin and depth of invasion of melanocytes [38, 70]. Although such information will not distinguish between malignant and benign lesions directly, it is nonetheless useful in differential diagnosis. Such techniques however, inherently rely on the calibrations of the colour image and even on the acquisition of supporting data as in the case of the light scattering model research which requires additional images taken in two infra-red bands. Diagnostic features based on absolute colour values are particularly sensitive to poor calibration, one such apparently effective indicator for melanoma was later found to be detecting nothing more than a capture process artifact [71]. The

requirements of this project to work with simple clinical images obtained from low-cost commercial video capture equipment, with calibration resting only on the assumption of subjective fidelity therefore all but precludes the use of spectral analysis techniques.

Colour data can also provide invaluable information in the identification of the lesion extent as the boundary between lesion and skin is often better defined in a colour image, especially where the lesion contains regions of light tan and grey which result in a similar grey-scale intensity to that of normal Caucasian skin. The colour information can be utilized directly with pixels being analysed in terms of their colour, or indirectly by the application of a pre-processing step manipulating the colour information to extract certain properties and provide an 'enhanced' image for processing. The accurate and robust identification of the lesion boundary is a vital element in the image-based analysis of skin lesions for both the diagnostic information it can supply directly, and in the assessment of lesion area features.

Identification from colour information of either irregularity in pigmentation, specific colours or the lesion extent can all be essentially reduced to the same image processing goal — the segmentation of the image into regions of distinct colour. It is with this goal in mind then, that the research presented below proceeds.

5.4 Previous Research in Colour Segmentation

The importance of colour in the diagnosis of skin cancer has naturally caused colour analysis to attract considerable attention from researchers in the field of image-based skin lesion analysis. The work falls into two categories with the emphasis on either the generation of a boundary for the whole lesion, or the detection of colour variation within the lesion.

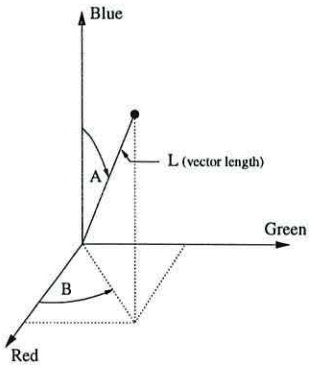
Dhawan *et al* [14] describe a colour and texture approach that aims to generate a boundary as well as segmentation within the lesions. They use a transformed colour–space which aims to extract the three components of intensity (I_1), coarse colour variation (I_2), and fine colour texture (I_3) which are defined as:

$$I_1 = \tfrac{1}{3}(R+G+B) \quad I_2 = R-B \quad I_3 = \tfrac{1}{2}(2G-R-B)$$

A multi-channel approach is then used; the intensity is processed using a pyramid segmentation scheme while Generalized Co-occurrence Matrices are used to evaluate the texture information contained in I_2 and I_3 planes. The information is merged with the intensity segmentation to form the final segmentation.

Dhawan *et al* [46] also describe the application of this method to images obtained using the Nevoscope. The segmentation data can then be combined with 3D information available via this device in order to improve prospects for accurate diagnosis.

Umbaugh *et al* have published a number of articles concerning colour processing of skin lesion images. The first of these [39] reviewed here concerns the detection of variegated colouring. The need for automated identification of features demonstrated by the AI/DERM [40] system is discussed together with the consequent demand for an effective method of colour segmentation. A detailed treatment of the background to colour representation is given and a number of colour-spaces are explained and discussed. The spherical coordinate transform (SCT) of RGB data is introduced which is defined as in figure 5.6.



$$L = \sqrt{R^2+G^2+B^2}$$

$$A = \cos^{-1}\left[\frac{B}{L}\right]$$

$$B = \cos^{-1}\left[\frac{R}{L\sin(A)}\right]$$

Figure 5.6: Definition of the SCT colour-space.

Segmentation is performed on only the A and B angle data, the L (brightness) was not used in order to avoid shadows causing regions to be split falsely. The maximum and minimum of A and B for the image are found and the two ranges thus defined are split into equal portions depending on the number of final 'colours' desired. Object filtering and labeling stages then remove small objects and identify connected colour areas. A simple test was used to convert the segmentation into

presence/absence decision on variegated colouring. The method was tested on a set of images classified by a dermatologist (cases identified as *possibly* showing variegated colour were excluded) and the technique was shown to be reasonably effective.

In the second paper, Umbaugh *et al* [12] present a border finding technique based on colour segmentation employing the Karhunen-Loeve or Principal Component Transform (PCT) together with an automatic colour quantization technique. The process begins by creating a sub-sampled version of the original for both data and noise reduction. This is achieved by taking the mean of each 8×8 pixel block to yield a 64×64 colour image. The PCT is then performed. The three components of the PCT space are experimentally found to contain approximately 91%, 6% and 3% respectively of the variance of lesion images. Segmentation is performed on the PCT space by the median-split method:

> The axis with the greatest variation is determined and the points are then divided into two equal sized groups (divided by the plane perpendicular to the axis which passes through the median value on the axis). This process is repeated until the desired number of output colours is reached. Averages are then calculated for each of the regions formed and each pixel is mapped to the nearest of these averages.

Results in relation to 'feature scores' made by dermatologists are compared for a number of original colour-spaces, with spherical coordinates found to yield the greatest success rate and RGB chromaticity to have the lowest error rate.

In another paper Umbaugh *et al* [72] this work is extended from binary segmentation (border finding) to colour *feature* identification. The performance of the PCT/median-split algorithm reviewed above is compared to the SCT/centre-split algorithm (four or nine colour regions). The detection performance on six features (tumour, crust, hair, scale, shiny and ulcer) forms the basis for the evaluation and is compared to feature maps produced by dermatologists. PCT/median-split results are given for six colour-spaces including the three described later in this chapter. The SCT/centre-split algorithm is only defined for the original RGB space. The results show that PCT/median-split performs best using the chromaticity space with $L^*u^*v^*$ having the second highest success rate. PCT/median-split on any of the six colour-spaces out-performed the SCT/centre-split method.

Ercal *et al* [13] focus on creating a boundary and use a transformation to reduce the RGB colour-space to a single component and then apply a threshold. They review a number of colour systems including the colour-space described by Dhawan above. The transformation they elect to use is $X = w_r R + w_b G + w_b B$ where the weights are calculated to maximize lesion/skin discrimination by comparing the colour character of two small regions extracted from the image, one from the lesion and the other from the surrounding skin. They show that the component plane X from this transform is consistently close to the first component obtained from the PCT, but that it requires less computation.

Hance *et al* [37] completed a comparison of a wide range of approaches for the generation of a border outlining the lesion. They find the PCT/median cut (described by Umbaugh and reviewed above) produces the most promising results, with adaptive thresholding and fuzzy c-means (algorithm described in detail in Lim and Lee [73]) showing potential. One of the methods they investigate, and which shows relatively poor performance, is the basic quad-tree split and merge algorithm, performed on the principal component of the colour data. Although part of the work presented in this chapter also uses a quad–tree it should be noted that this structure is only used in the early stages of the technique, with the majority of the processing being performed on a Region Adjacency Graph (RAG) constructed from the quad–tree.

Colour based image segmentation naturally attracts research interest beyond the analysis of skin lesion images. A particularly interesting approach to natural scene segmentation is presented by Panjwani et al [74] who propose a general colour–texture segmentation method using region merging with uniformity judged through Gaussian Markov Random Field Models. The region-based segmentation technique presented later in this chapter is similar to this method in that both methods use splitting followed by an agglomerative clustering phase where this consists of conservative followed by optimal merging on a RAG. However, amongst other things,the mechanics of the uniformity measures and the conservative merging stage used differ from those presented here.

The use of ELM images in colour analysis of skin lesions is common as ELM image acquisition avoids the specular reflections from the stratum corneum which

plague normal clinical images. Specular reflections effectively result in anomalous data disrupting otherwise homogeneous regions of colour so that ELM images are 'cleaner' and therefore more amenable to colour-based segmentation. It is also important to realize however that images acquired using ELM are fundamentally different in character to normal clinical images; By rendering the skin surface semi-translucent, a whole new set of features is revealed. The differences between the two imaging methods mean that techniques which are effective for colour segmentation and feature extraction in ELM images are not necessarily useful for clinical images, and further, conclusions concerning the performance of techniques and diagnoses are certainly not transferable. With such considerations in mind however, the techniques and insights gained from such research may still be useful.

In general, the colour segmentation techniques which have been previously used for lesion image analysis can be seen to consider each pixel in the image as an individual entity and often require the number of output 'colours' to be specified in advance. Both of these features are in conflict with the aims of lesion image segmentation as discussed in the following section.

5.5 Segmentation of Colour Images

This section explores the aims and goals underlying the segmentation of skin lesion images based on colour information. A clear understanding of the elements that constitute good performance in a segmentation outcome is essential for the development of any useful scheme for lesion analysis.

The goal of colour based segmentation is to divide the image into areas of homogeneous colour. The precise meaning of this statement is however unclear on two counts: Neither the term 'areas' nor what is meant by 'homogeneous colour' is well defined. The derived system concept for this project requires that the indicators used to construct a diagnosis are amenable to explanation to the patient and visual verification. In terms of the segmentation of lesion images, and especially given these project requirements, it is clear that the value of a segmentation

relates directly to human perception. The following two sections discuss each of the two terms in relation to this measure of value.

5.5.1 Regions

Given an image of a lesion it is clear that a statement that the lesion contains 'different colours' is intended to convey the situation where the lesion is composed of a number of areas that have internally consistent, but mutually different colour. The colour of the individual pixels is clearly not what is truly important, for example, scattered but isolated white pixels in the lesion area would not equate to the statement that the lesion contains areas of white. Such observations reveal that the term 'areas' is, in this application, better expressed as 'regions'. A useful segmentation of a lesion image should then aim to show local connected regions of homogeneous colour.

Many approaches to image segmentation, such as K-means [75] and standard neural network implementations [76] as well as most of the techniques investigated in the literature as outlined above, follow the pattern of classification of each individual pixel as a separate entity. This view of the image data is clearly unhelpful since the lack of consideration of locality between pixels runs against the aim of reflecting human perception of colour regions. Individual classification is therefore far from ideal. Some of the problems of loss of locality information can be offset by the inclusion of smoothing prior to classification; by making pixels more similar to their neighbours, classes that are more spatially coherent are obtained and this in turn results in segmentations which are easier to interpret visually. Such processing is however undesirable due to the blurring effect at boundaries caused by the combination of different colours.

Furthermore, the majority of the techniques reported in the literature also require the *advance* specification of the number of 'colours' to divide the image into. This situation is obviously counter-productive since the variegated colouring differential indicator inherently entails that melanoma images are likely to have many more different 'colours' than benign moles. Consequently a number of 'colours' large enough to perform well with such melanomas will cause false

'colour' divisions in a benign mole.

A process which embodies the idea of locality in segmentation, and which divides the image in relation to a given degree of colour difference, is therefore most suited to this task.

5.5.2 Colour Similarity

The second term needing clarification in the statement of the goal of lesion image segmentation is that of 'homogeneous colour'. When the aim is to divide the image based on differences in colour it is vital to identify what exactly is different about the colours that should be separated and similar for those that make up homogeneous regions.

Much of the work that has been done in colour analysis of skin lesions has used the RGB colour-space, and given the ease in obtaining data in this form from the type of commercial camera envisaged for this project, it seems sensible to begin by examining this colour-space.

In many cases this three dimensional colour-space is immediately processed with the aim of reducing the colour information to a single dimension to simplify processing. This is often achieved by taking the actual, or an equivalent to, the first dimension after the application of the PCT, in other words the first principal component (PCT_1) [13, 37]. The effect of the PCT can be viewed as finding a new basis for a multi-dimensional space in which the new basis vectors are aligned with respect to the greatest variance in the data. In the new basis, the dimensions (or components) are selected *in order* so as to reflect the greatest variance in the remaining dimensionality of the space given all the previously selected basis vectors. The selection of PCT_1 then reflects the projection of the RGB colour data points onto the line of greatest variance in the data as a whole. This action can be justified through the observation that the RGB data for a skin lesion image normally has an obvious variance maximum orientation and little variance outside of this axis so that the majority of the information content is reflected in PCT_1.

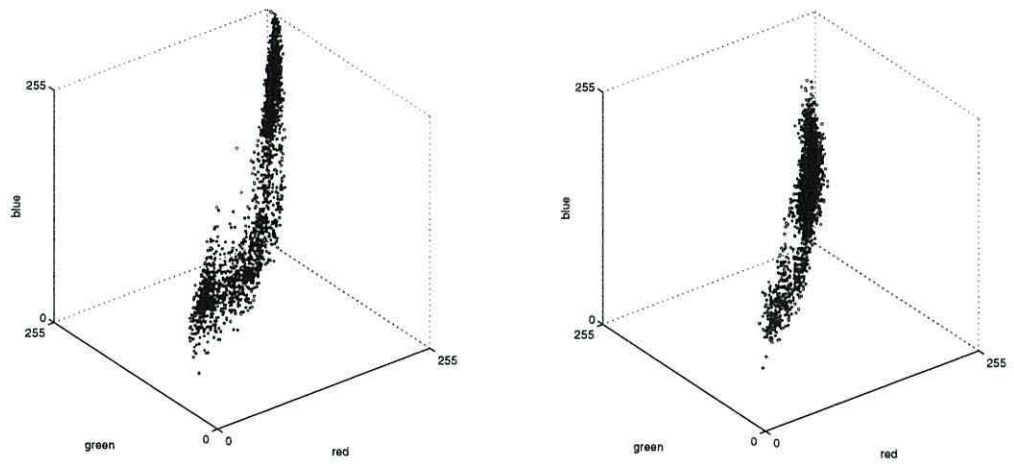


Figure 5.7: RGB colour-space scatter graphs for two images

An analysis of the distribution of the RGB colour points for several lesions has demonstrated however that this is perhaps a mistaken simplification as the PCT_1 axis is often surprisingly close to the intensity axis, so that the result is little different from processing a grey-scale version of the image. Furthermore it is clear that a single dimension obtained from any linear transform of the RGB space would be unable to capture all the main modes of variation found in the image data as the data generally occupies a distinctly curved lozenge-shaped subspace. Figure 5.7 shows representations of the RGB colour-space for two example lesion images. The orientation of greatest variance is clear in both cases as is the fact that the PCT_1 will be close to the intensity axis and the curved nature of the subspace occupied is also readily apparent. These observations indicate then that the analysis of PCT_1 alone is perhaps a mistaken simplification as the evidence shows that much of the *colour* information is being discarded.

Reducing the dimensionality of the colour information has many problems as identified above. These problems prompted the development of a region-based technique for the segmentation using the original full three dimensional RGB data.

Whatever the representation chosen for colour information the need to detect regions of 'homogeneous colour' requires a method for assessing the similarity of two colours as represented in that space. Perhaps the most obvious measure for a simple space such as RGB is the *Euclidean distance* or L2 distance which is defined as, $\text{L2}(p,q) = \sqrt{\sum_i (p_i - q_i)^2}$. This measure is both widely used and well

understood and quantifies the actual distance between two points in a Euclidean space such as RGB.

In summary, section 5.5 has examined in detail the issues and underlying goals of segmentation for lesion image analysis. The importance of a clear definition of the terms in the phrase 'areas of uniform colour' was exposed. The term 'areas' has been seen to entail a need for locality consideration when classifying pixels. Consideration of colour similarity has demonstrated that common dimensionality reducing preprocessing does in fact discard much of the *colour* information.

5.6 Multi–stage Region Agglomerative Clustering (mRAC)

This section describes a novel agglomerative clustering technique, mRAC [77], developed in response to the identified need for a region-based approach to colour segmentation of skin lesions.

Clustering methods aim to partition data into groups consisting of elements which possess similarity. The mRAC process has two criteria for similarity; the first is explicit and is explained below, the second is implicit - two clusters must be spatially adjacent to be considered uniform. This is not usually considered to be a uniformity criterion, however it is an important defining influence on the character of the final clustering solution produced. Here, clusters will be called regions.

5.6.1 The mRAC Algorithm

This algorithm is similar in some respects to that described in Panjwani et al [74] which presents a general colour–texture segmentation method using region merging with uniformity judged through Gaussian Markov Random Field Models. Both methods use splitting followed by an agglomerative clustering phase where this consists of conservative followed by optimal merging on a Region Ad-

jacency Graph (RAG). However, the mechanics of the uniformity measures and the conservative merging stage used differ from those presented here.

The uniformity criterion used relates to the Euclidean distance d between the three vector means (mean colours) of two regions i and j and their union. Two regions are "uniform" and will be merged if d is small. The value d will be referred to as the "distance" between the regions, and is given by:

$$d(i,j) = \max\{|\mu_i - \mu_j|, |\mu_i - \mu_k|, |\mu_j - \mu_k|\} \quad (5.17)$$

$$\text{where} \quad k = \{i \cup j\}$$

$$\text{and} \quad \mu_i = [\mu_{i_1} \ldots \mu_{i_n}] \quad \text{is the vector mean for region } i$$

The process begins splitting the image into small regions which will be joined together as the clustering proceeds. This is achieved by creating an extended quad–tree representation of the image, the extension being that where a region is not exactly divisible into four sub–regions, the top left sub–region (at every level in the tree) is allowed to be larger by one pixel in either direction. This allows the representation to proceed down to regions of the smallest possible size (either width or height of 1 pixel) regardless of the original image size.

Splitting for the initialisation continues to one level before the limit, i.e. when a region has a smallest side length of just two pixels. Each region is then tested to ensure that its vector mean (colour mean) is similar to that of each of the four sub regions it contains. If any one difference is above a small threshold t_s then the region is split. This is superior to the simple approach of taking the initialisation right to the limit (dividing the image into the smallest possible regions) as it produces a smaller number of regions to be considered whilst ensuring representation is equally good. A good representation cannot be guaranteed if splitting begins at a higher level as very small features might not be sufficiently significant to cause a large and generally uniform region to be split at that level.

Once this step has been completed a RAG is constructed which lists all the connections between regions, where a region is considered to be connected to another

if it shares a common boundary (in a 4-connected sense). This RAG is then the basis of an agglomerative clustering process that forms the final segmentation.

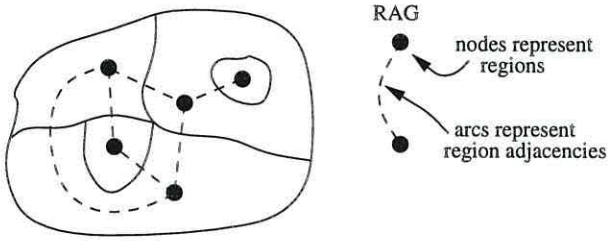


Figure 5.8: Regions and a Region Adjacency Graph

The conservative merging phase is necessary to reduce the number of regions and hence the number of candidate merges in order that the optimal merging phase becomes computationally realistic.

Each adjacency is tested for uniformity using the same measure as above. For every adjacency where the distance between vector mean (colour mean) is less than a threshold the regions are merged. The phase is executed in three passes with the threshold increasing in even steps up to t_{cm}, this ensures early region formation based on strict similarity, allowing the means to stabilize. Without this step-wise increase many merges of tiny regions that have differences close to the threshold can cause a large drift in the mean and the formation of a non-uniform region.

In the optimal merging phase the same test is used, however this time the process is iterative and only the single best possible merge is performed. The best merge is defined as the joining of the two adjacent regions k and l whose vector means (colour means) are closest, given by,

$$\begin{aligned} &\text{let A be the set of all adjacent region pairs, then} \\ &\qquad\qquad \text{optimal merge is } (k,l) \text{ such that} \\ &\qquad\qquad d(k,l) = \min_{i,j \in A} d(i,j). \end{aligned} \tag{5.18}$$

The step-wise approach is necessary as each iteration alters both the list of possible merges and the distances associated with them (the new combined region having a new vector mean). The optimal merging phase stops when the best possible merge has a distance which exceeds a threshold, t_{optm}.

Finally, the segmentation is cleaned when the distance for the optimal merge nears t_{optm}. This cleaning involves forcing the merge of all very small regions (less than 0.1%, say, of the image area) to the adjacent region that is most similar. The premise for this forced merging is that any small region that has not yet merged must be an anomaly in the data. The addition of this phase makes the final segmentation easier to interpret.

5.7 Evaluating mRAC

In order to assess the potential of mRAC for the segmentation of lesion images, the method was applied to a set of example images and the performance evaluated in terms of visual quality of colour separation.

The images used in the trials of the technique [3] include examples of several types of lesion including malignant melanoma and compound naevi. The original images are 24-bit full colour digitized from slides with approximately 4 micron pixels. Each image is sub-sampled to produce 350x230 source images.

All parameters were kept constant for all the results presented in this section. The exact parameter settings are not critical, different but similar parameter sets generate results differing in character but not in validity. The setting of t_{optm} obviously has the greatest effect on the final segmentation. The thresholds used were; splitting $t_s \approx 7.5$, conservative merge $t_{cm} \approx 10$ and optimal merge $t_{optm} \approx 38$. These thresholds relate to the expected range of the uniformity criterion d which theoretically (for the 24bit 3-element colour data) had a range of $0 - \sqrt{3 \times (255)^2} (\approx 440)$. The segmentation was cleaned when the optimal merge phase reached 90% of the t_{optm} threshold for regions less than 0.15% of the image area.

The fixed thresholds used here are valuable in that they ensure comparability between the results for different lesion images. Regardless of the properties of the

[3] The images were provided by V. Wallace and Dr. J. Bamber of the Physics Department, Clinical Research Centre, Royal Marsden Hospital, Sutton, permission to use them is gratefully acknowledged.

current image being analysed, the same degree of difference between the colours of two regions should be required for the regions to be considered distinct. This is important as it would be undesirable to label two regions of similar colour as distance simply because the lesion as a whole shows little variation, as would be the result of a variable or adaptive methods for determining thresholds.

Results are presented for five images, three containing a malignant melanoma MM and two containing a benign lesion (Compound/Blue Naevus). These examples, especially the malignant melanoma images, were selected from the test set to best illustrate the performance of the method, having a rich variation of pigmentation or showing other colour features.

5.7.1 Results

Figure 5.9 shows the segmentation performance which can be achieved through this method for the five example images, the green lines show the boundaries between the identified colour regions. The third example in the figure shows an unusual failure where the colour mean of much of the lesion area is not sufficiently different to that of the surrounding skin to produce a good result.

The mRAC method shows clear potential for the detection of regions of differing pigmentation colour in images of malignant melanoma using RGB data. This information could be used together with a boundary detection method to provide an indicator for differential diagnosis. The method could also provide useful lesion/skin segmentation data which could be used to support the validity of a previously acquired boundary segmentation.

In some cases regions of apparently different colour have been unexpectedly merged into a single region. This indicates that the Euclidean distance similarity measure on RGB triples is perhaps not the best measure for colour uniformity in lesions. As noted previously, lesion images consist mostly of variations through white—red—brown—black which is a narrow band diagonally through the RGB colour–space. This necessarily limits the discriminative power of the Euclidean distance measure.

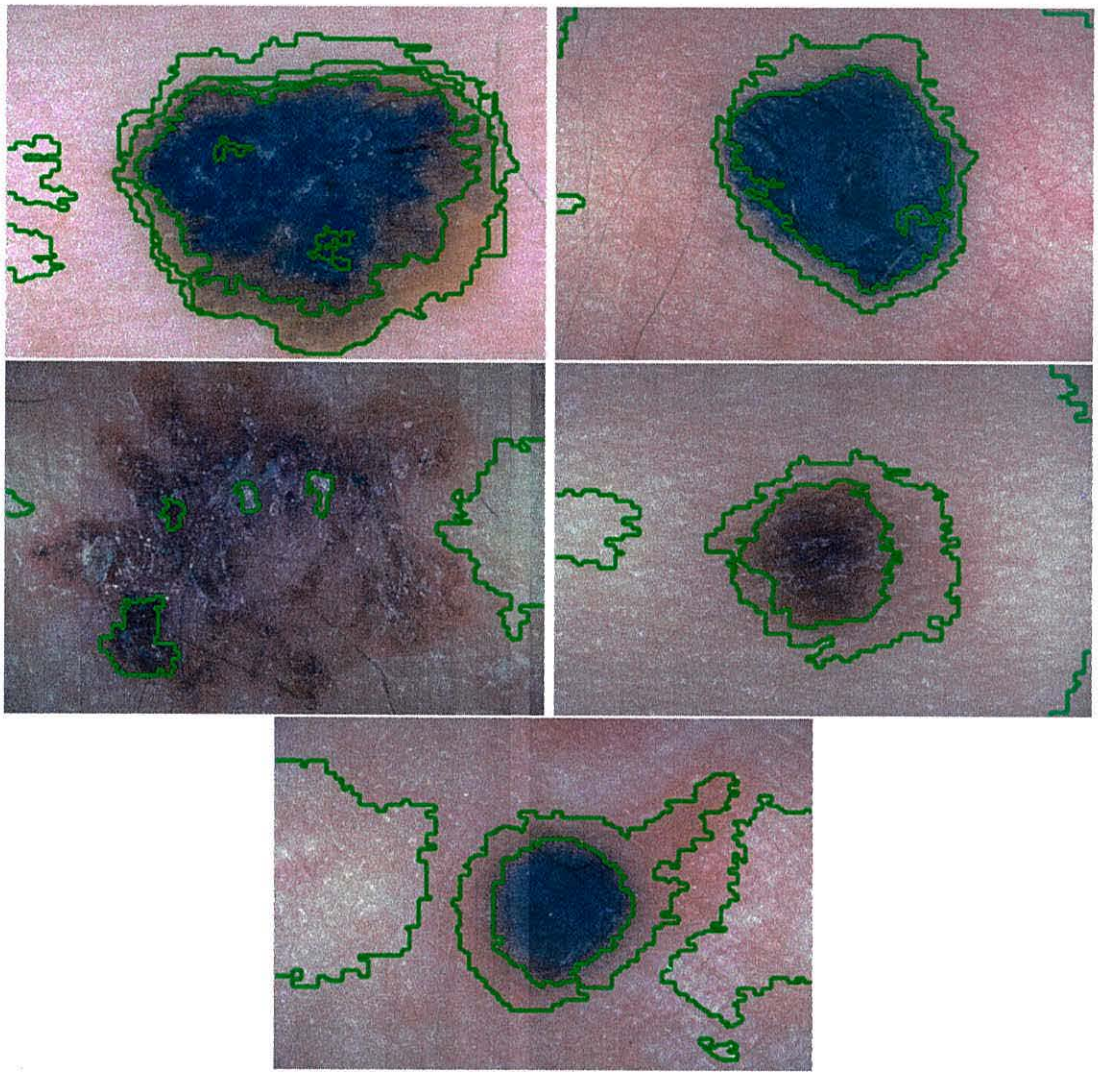


Figure 5.9: Examples of mRAC segmentation using RGB data: Left to right and top to bottom: MM with colour variation in surrounding border, non–MM with little colour variation, MM with many patches of differing colour, MM with colour variation in surrounding skin, and non–MM with red inflamed surrounding skin.

5.8 Colour-spaces and Colour Difference Metrics

The results of the initial trial of the mRAC region based classifier were promising in the segmentation of the lesion images. It is however apparent that there are some considerable differences between the colour grouping which results from the use of the L2 distance measure on the raw RGB data and that of the human observer, especially in the area which comprises the transition from lesion to skin.

The conclusions drawn from the RGB mRAC trials highlight the problems caused by the presence of only a small subset of the full colour range in lesion images. The segmentation performance obtained with the standard L2 measure is significantly reduced when applied to this lozenge-shaped subspace. Such difficulties prompted the investigation of alternative colour-spaces and colour distance measures with the aim of better reflecting colour differences as perceived by the human observer.

5.8.1 Colour-space Use

The goal of colour based segmentation, as identified earlier, is to divide an image into areas of homogeneous colour. In order to obtain a good segmentation, perceptually similar colours should form identifiable clusters in the colour-space. Perhaps the most intuitive interpretation of this statement is that of a spatial cluster for each perceptual 'colour' where the colour data is viewed as points in a three dimensional Euclidean space. The actual clustering desired for a certain application could of course correspond to other cluster forms in the colour-space, for example as ellipsoids or even concentric spheres. It is vital then to examine the distribution of the colour data within the colour-spaces in order to identify cluster forms which would result in the properties desired for the final segmentation. In the domain of lesion image analysis for example, it is desirable for the entire skin area to be classified as a single region, with the lesion divided into one or more regions of differing colour.

This section examines the colour distributions for lesion images in the three colour-spaces identified in section 5.2. for each colour-space several example images are shown as 3D scatter plots with red points representing the colours in the lesion area and blue points representing the skin. For IHS and $L^*u^*v^*$ 2D chromaticity plane plots are also presented in which the points are shown in their original colour. The colour images were sub-sampled to 20% of their original size to obtain a manageable number of points.

RGB

The distribution of colour data in the RGB space has already been touched on in section 5.5.2. Figure 5.10 shows two typical scatters in the RGB space. The skin

colours clearly occupy an extended region approximately following the intensity axis, corresponding to a range of relatively high intensity colours from light tan to white-pinks. The lesion is represented by a range of darker colours extending from dark browns (which occupy a region extending on the same axis as the skin colours) to blue-blacks. The blue-black colours are not seen in all lesions, and correspond to the presence of melanin in the dermis [38]. Although the blue-black colours are a common feature of melanoma, this is by no means an exclusive property; Blue naevi, for example, derives its name from this blue colour.

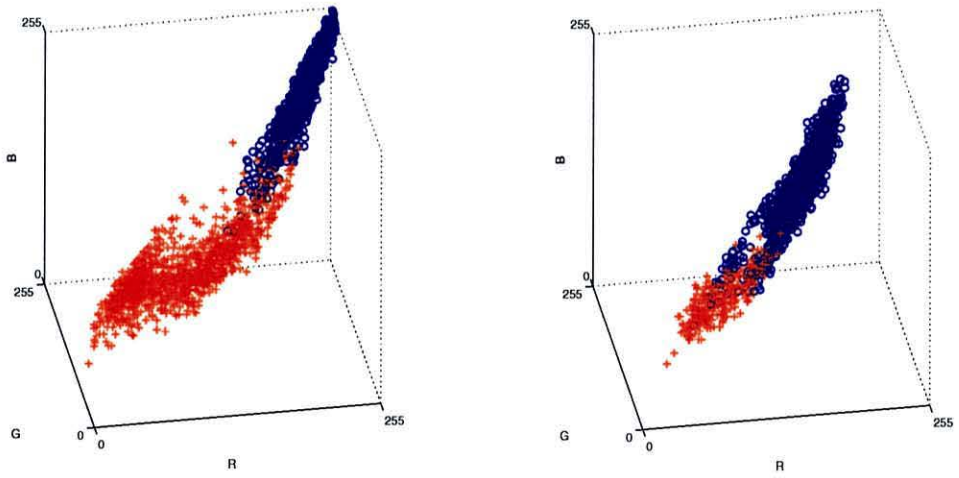


Figure 5.10: RGB colour-space scatters for two images. Red points are from the lesion area and blue points are from the skin.

IHS

Consistent with the underlying motivation for this colour-space, the distribution colour lesion image data is simpler to explain in this space. Figure 5.11 shows scatters for the same two typical images as used above, with the chromaticity plane horizontal and intensity as the vertical together with hue-saturation plane views (effectively looking at the 3D scatter from above) where the points are shown in their original colour. In the 2D hue-saturation space, the skin colours are clearly seen as a compact group just to the right and slightly above the achromatic centre of the plane. These white-pink skin tones appear as the vertical spike in the 3D view showing that the skin consists of varying intensities of a single colour. The darker brown lesion colours can be seen to be an extension of this same chromaticity (with an increasing variation (more often an increase in saturation i.e. further right and up from the achromatic point.) The blue-black lesion colour

(of which there is little in the right-hand example) is seen as a long tongue away from this vertical skin-and-brown-lesion region. In the chromaticity plane, the blue-black lesion colours extend to the left and slightly below the achromatic point.

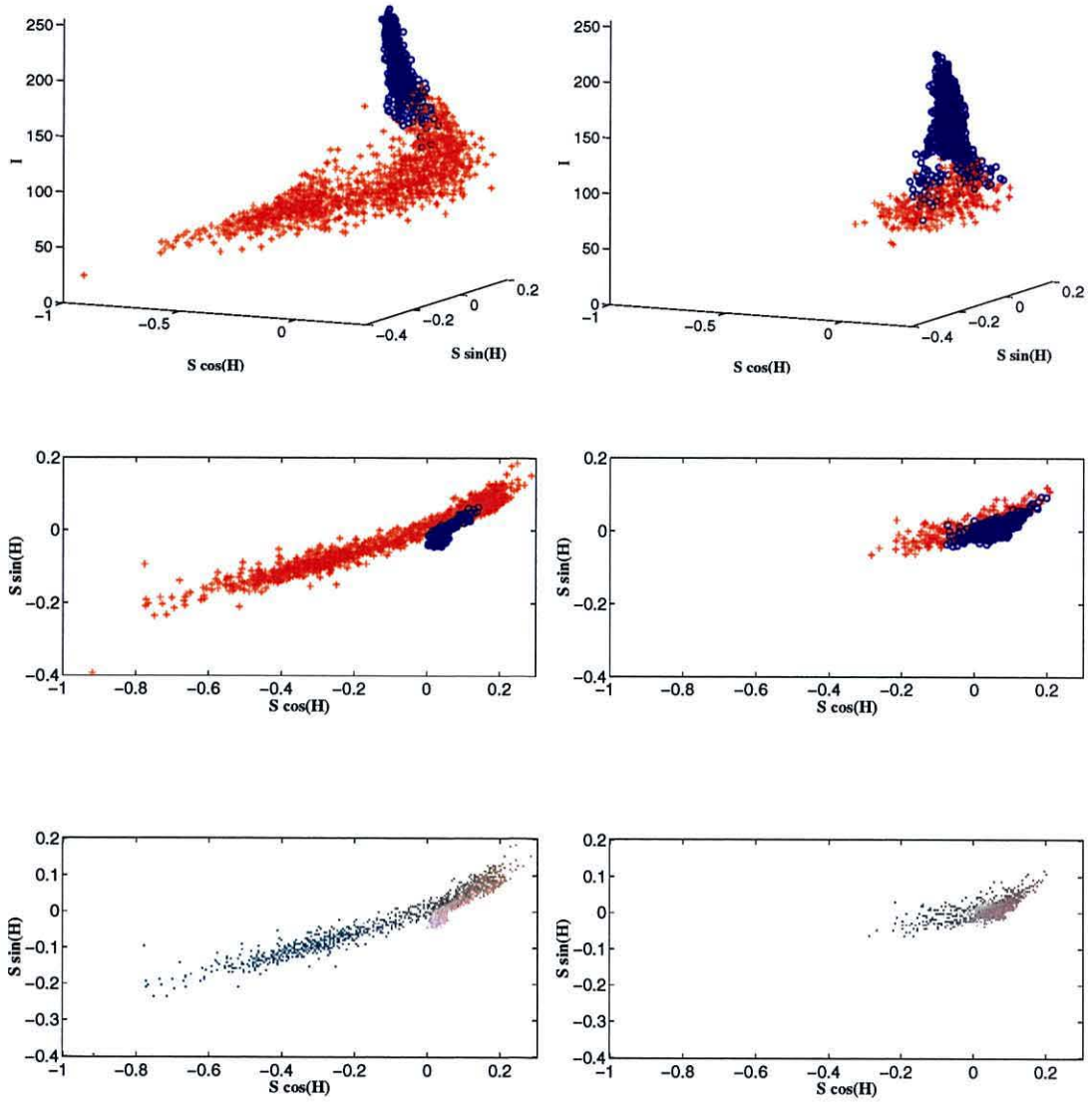


Figure 5.11: IHS colour-space views for two images: Top and centre, a 3D scatter representation and a hue - saturation plane view respectively where red points are from the lesion area and blue points are from the skin. Bottom, a second hue - saturation plane view where the points are shown in their original colours.

CIE $L^*u^*v^*$

The spatial distribution of the colour points in the $L^*u^*v^*$ colour-space has many features in common with the IHS space. Both are based on the separation of the chromaticity from brightness and it is this common underlying concept which results in the similarities seen. Figure 5.11 shows scatters for the same images as before. As with the IHS data, the chromaticity plane is the 'horizontal' plane of the 3D projections and luminosity is shown as the vertical. The chromaticity plane is again shown separately as a 2D plot in which the points are shown in their original colour.

The most striking differences between the IHS and $L^*u^*v^*$ data is the way that the skin colours have been collapsed from the extended lozenge-shaped region to one much more closely resembling a sphere. Since $L^*u^*v^*$ is intended to be a 'uniform' colour-space the spatial relationships in the 3D scatter of the different lesion colours should in general be seen as spherical clusters. It is also apparent that the constituent colours of the lesion image are generally more spread out in $L^*u^*v^*$ than in IHS, with the skin and lesion points showing a much clearer separation (seen clearly in the right-hand example). The long blue-black tail in the left-hand example is also noticeably less pointed.

In conclusion then, the IHS space presents a platform through which it is much easier to identify and describe the clusters that correspond to the constituent 'colours' of the lesion images than in RGB. The $L^*u^*v^*$ colour-space presents the data in a similar way to the IHS space (they are both brightness-chromaticity decoupled spaces) and seems to retain all the advantages over RGB offered by IHS. Moreover $L^*u^*v^*$ shows a better separation of the identified colour clusters, and perhaps most importantly, the white-pink-tan skin tones are concentrated into a single near-spherical region in the space. The value of this last property being that the skin is generally seen as having a single colour (human observers rarely identify more than one colour in the skin area).

The IHS space, when compared to RGB undoubtedly offers a platform from which more effective and useful colour segmentation of the lesion image data could be obtained. However, an effective colour difference measure for this application would have to be developed which could account for the non-spherical clusters

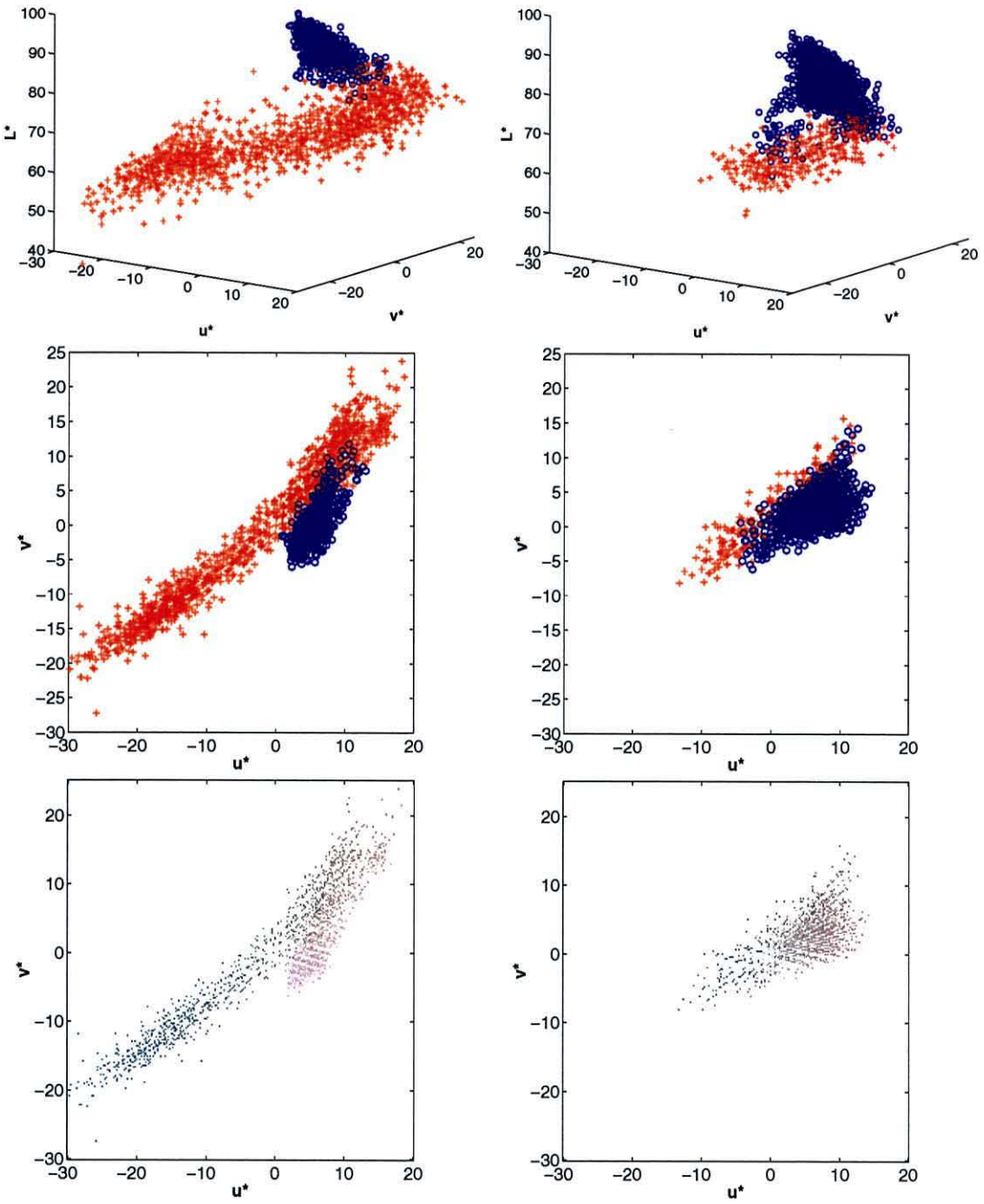


Figure 5.12: $L^*u^*v^*$ colour-space views for two images: Top and centre, a 3D scatter representation and a u^*v^* plane view respectively where red points are from the lesion area and blue points are from the skin. Bottom, u^*v^* plane views where the points are shown in their original colours.

formed in this space. The advantages offered by the $L^*u^*v^*$ colour-space identified above when taken together with its basis on a simple pre-defined measure of

colour similarity known to approximate human perception, clearly indicate that this space should provide an effective starting point for segmentation to identify the regions of different colour in a lesion image.

5.9 Evaluating the mRAC—$L^*u^*v^*$ Combination

The investigation of colour-space representations and colour difference metrics demonstrated the suitability of the $L^*u^*v^*$ space and its associated L2 difference measure for the colour analysis of skin lesion images. After transformation of the colour data to the $L^*u^*v^*$ space the mRAC region-based segmentation technique was applied.

The new data requires a reconsideration of the parameter set for mRAC. A single set of parameters was desired (and used) for all the sample images; The advantages of fixed threshold parameters being identified in the processing of the original RGB data with mRAC. In the $L^*u^*v^*$ colour-space the 'component colours' have been seen to be, at least to a certain extent, identifiable as spherical clusters in the 3D projection of the space (consistent with the 'uniform' nature of the $L^*u^*v^*$ colour-space under the L2 distance measure). In particular the 'skin colour' is generally identifiable as a single distinct spherical cluster. Since this skin colour is normally perceived as a single 'colour' it is reasonable to obtain a first estimate for the final stopping threshold t_{optm} so as to include the skin cluster as a single colour region. In general the skin cluster forms a near-sphere with a diameter around 15 units (for 0-255 original RGB data), as can be confirmed by considering the scatters in figure 5.12. In practice, a lower t_{optm} was found to be more effective in separating the colours perceived as different in the lesion area. The setting of t_{optm} has a much greater effect than that of t_s and t_{cm}. t_s should be set so as to ensure that only genuinely uniform regions remain un-split at that stage and t_{cm} so as to reduce the number of regions to a computationally manageable number for the final stage. The thresholds used here are; splitting $t_s \approx 2.5$, conservative merge $t_{cm} \approx 3$ and optimal merge $t_{optm} \approx 9$. The segmentation is cleaned when the optimal merging phase is complete – any regions less than 0.15% of the image area are force-merged into their most similar neighbour.

5.9.1 Results

Results are presented for the same five images used in the mRAC–RGB test (figure 5.9) for comparison purposes although a much larger number of images have been processed. Figure 5.13 shows these examples of the segmentation performance which can be achieved through this method. In general, and as expected the segmentation corresponds much better to human perception of the different colours and their extent than the RGB based results. Examples of clearly improved performance are:

In the top-left image: The upper limit of the lesion regions now follows closely the boundary between the brown lesion and the skin.
The central blue-black region is now identified as a different colour.
The orange-brown outermost parts of the lesion and the darker brown regions within this are now identified separately along a division that agrees well with visual inspection.

In the top-right image: The lesion border regions follow more closely the edge of the lesion.

In the middle-right image: The skin is now a single region and the lesion region extents correspond better to the limits of the two colours there.

In the bottom image: Again, the skin now falls into a single region and is not divided by the brighter areas at either side.
The different colours surrounding the central blue-black region are detected.
The red inflamed region extent is more clearly matched and the similar red region at the top of the image is detected.

The segmentation performance on the middle-left example has however shown little improvement with the new process. In fact the two methods produce segmentations with surprisingly similar properties – the regions at the bottom left and right-hand edge identify the same image features. It is clear that the original image in this case is much darker than the others in the example set and there is little brightness range in the image. This situation could well be the result of a

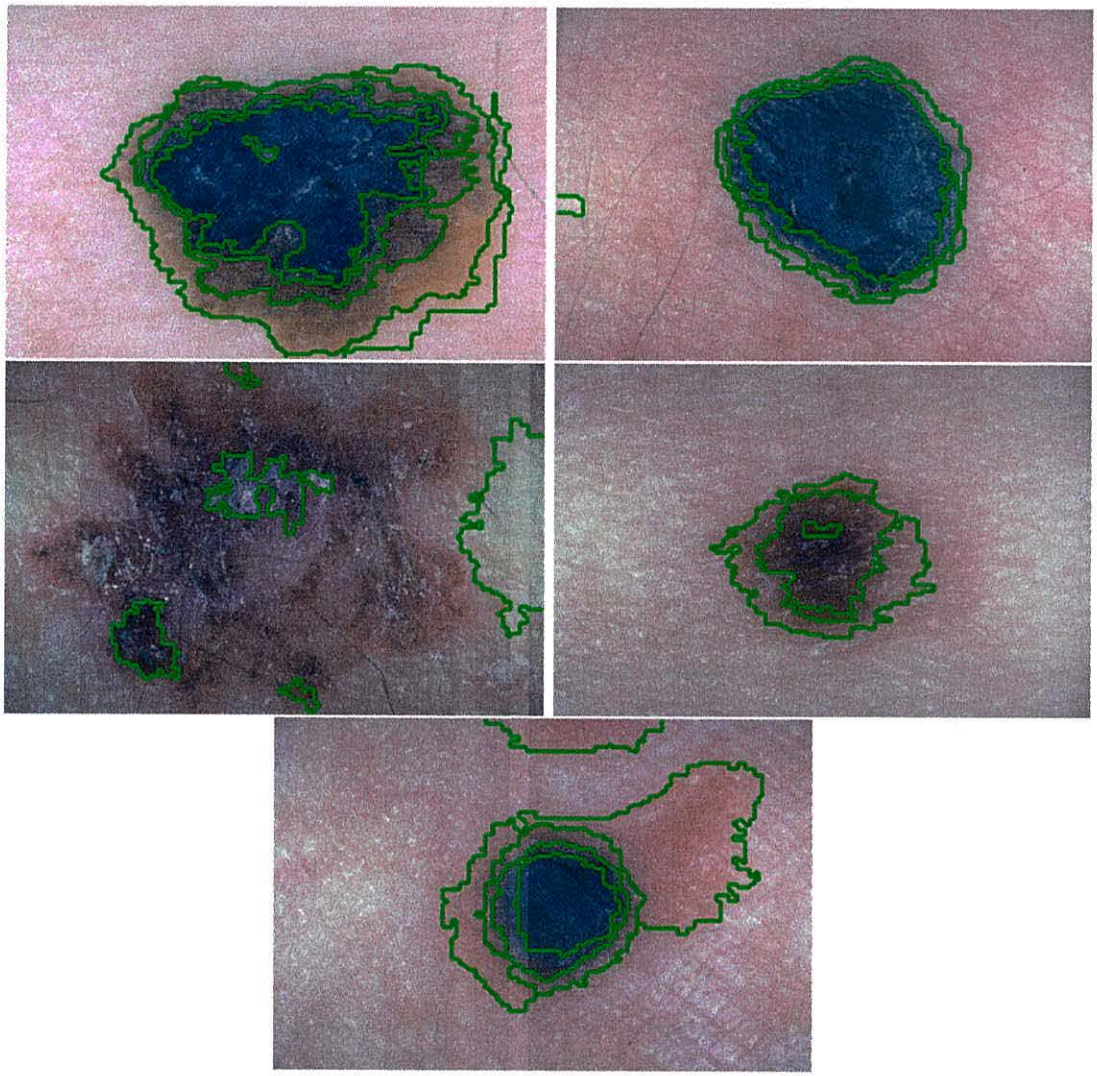


Figure 5.13: Examples of mRAC segmentation using $L^*u^*v^*$ data: Left to right and top to bottom: MM with colour variation in surrounding border, non–MM with little colour variation, MM with many patches of differing colour, MM with colour variation in surrounding skin, and non–MM with red inflamed surrounding skin.

poor acquisition of the actual scene, in which case the use of a simple calibration card in-shot would provide a means to avoid this poor performance.

An examination of the colour-space use for this particular image (figure 5.14 left) indicates clearly why such problems have arisen with this particular example; All the data is concentrated in a single region in the space, with little variation in either luminosity or chromaticity. In spite of this, a useful segmentation can be obtained simply by reducing the final threshold t_{optm} to 6 (figure 5.14 right). Such a segmentation cannot be obtained using the RGB based processing as, when the

thresholds are set sufficiently low to ensure a good treatment of the lesion area, the overall segmentation becomes confused with many regions representing the skin area. This shows a further advantage of the $L^*u^*v^*$ colour-space over RGB for lesion image segmentation.

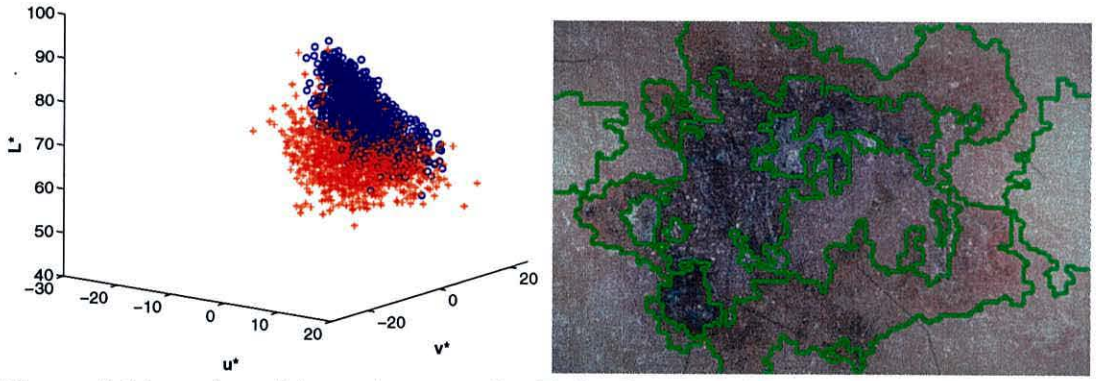


Figure 5.14: A problematic example: Left, $L^*u^*v^*$ colour-space scatter showing unusually limited variation, and right, a good segmentation obtained by reducing t_{optm}.

It is important to note that a potential problem exists in the assessment of segmentation performance by the simple means of visual inspection; Visual assessment of performance tends to prefer clean segmentation (one with few regions). Simpler results are often initially seen as better than the more complex solutions which upon detailed inspection are often seen to more faithfully represent the different colours in the image. Naturally, the more divided the image is, the better the separation of colours will be, however the differences found in the course of this investigation seem to go beyond this simple explanation.

The level of division required in the final segmentation is obviously directly affected by the aims and implementation of the process which is to use the data. For example, a binary division of the image based on colour would provide useful data for skin-lesion boundary confidence checking whereas a relatively fine division would be needed to investigate variegation in colouring. The setting of t_{optm} must then be considered in the light of the goals for the segmentation data.

5.10 Conclusions

Colour, a simple concept in essence, is found to be a complex concept when examined in detail. However, colour can be usefully (and much more simply) expressed and manipulated in terms of combinations of three fixed primaries; The RGB system, for example, forms the conceptual basis for much of current image capture, communication and display technology, with the red, green and blue primaries used having significance in relation to human vision.

Colour information is widely recognized as an important diagnostic feature in the analysis of skin lesions, and this is reflected in its prominence in the skin cancer identification checklists. The colour-based indicators used mainly relate to either the variegation of colouring or the presence of specific colours. Research into colour-based lesion diagnosis falls into the three categories of spectral analysis of accurately calibrated reflectance spectra, analysis of spectral properties of image-based information, and finally, the analysis of colour images. The envisaged primary-care diagnosis support system (considering its demands for low capital and running cost, minimal training for operation and explainable diagnosis) have already been shown to indicate an image based system. These requirements suggest the use of techniques only of the third category since special equipment is required by the first and the second demands not only accurate calibration, but often additional supplementary data.

Colour image processing for skin lesion images involves identification of either irregularity in pigmentation, specific colours, or the lesion extent. All these goals essentially reduce to segmentation of the image into regions of distinct colour. Due to the importance of colour as an indicator for skin cancer there has been considerable interest in colour-based segmentation for lesion images aimed at either finding a binary lesion/skin segmentation or detecting variegated colouring. The reported methods generally begin the analysis by reducing the dimensionality of the colour data, often to a single dimension oriented either in relation to maximum data variation or to maximize the distinction between identified lesion and skin sample regions. A variety of segmentation techniques have been employed on this transformed data with varying degrees of success.

The underlying goals of segmentation for lesion image analysis were examined in detail and the importance of a clear definition of the terms in the phrase 'areas of uniform colour' was exposed. The majority of the segmentation methods which have been used on lesion images consider the pixels as individual entities ignoring their location in the image, and require the number of output 'colours' to be specified in advance. The former is contrary to the goal of identifying contiguous regions in that the spatial position of the pixels is not being considered, whilst the latter is obviously problematic since variegated colour is a known differential indicator. The notion of uniform colour is usually related to proximity between colour data points in the colour-space. Consideration of the distribution in RGB space for lesion images indicates that mapping to a single dimension based on maximized variation (such as obtained from PCT_1) would discard the majority of the colour information.

The demand for consideration of spatial information prompted the development of a region-based approach to lesion image segmentation. The mRAC agglomerative clustering technique was developed and shown to be effective in identifying regions of homogeneous colour using the full RGB data and the L2 distance measure for colour similarity.

Although the results from the mRAC–RGB–L2 trials were promising, it was apparent that there were considerable differences between the colour grouping given by L2 on RGB data and that of the human observer. Many other colour-spaces have been used in image processing to satisfy the requirements of a wide variety of applications. Three colour-spaces were seen to be particularly relevant to the need for approximation to human perception, RGB, IHS and $L^*u^*v^*$. An investigation the general properties of and the distribution of the colour information from lesion images in these spaces clearly indicated the value of $L^*u^*v^*$ with its decoupling of luminosity and chromaticity information and its construction on the principle of consistency between L2 distance in the space and perceived colour difference.

The final results show the distinct improvement in segmentation performance obtained using mRAC with the $L^*u^*v^*$–L2 combination of colour-space and colour difference measure. Clear examples of better visual quality of segmentation are

seen in all but one of the presented test image results. The desire to ensure consistency in what is considered to be a 'different colour' demands the use of the same segmentation process parameter set for all images. The use of a single parameter set prevents the technique from producing a satisfactory result in all cases, however lowering just the final t_{optm} parameter, the $L^*u^*v^*$–L2 combination is able able to produce a good segmentation even in the difficult (dark and low contrast) example. No such solution can be obtained using the RGB–L2 combination as separation of the apparent colours in the lesion is only maintained in a highly complicated and confused segmentation.

The promising results obtained with the new method suggest that the segmentation information should be used not only for the quantification of variegation and presence of lesion colours, but also to provide support in boundary identification. However, the development required to convert the multi-region segmentation to a binary lesion/skin division and for investigation of combination methods and confidence rating using multiple boundary estimates could not be addressed in the time available.

In summary, the importance of colour as an indicator in skin lesion analysis requires its consideration for inclusion in any feature set for a diagnosis support system. The requirements of the particular low-cost system envisaged limit such analysis to processing of a standard colour image with calibration based only on visually confirmed fidelity between the captured image and scene. The analysis of such images requires their segmentation into regions of perceptually uniform colour. The requirements of such segmentation have been analysed and a technique is presented which performs well on the majority of the sample images used in this study.

Chapter 6

Texture – Skin Pattern Modelling

This chapter examines the possibilities for obtaining diagnostic feature information through the study of texture data. The nature and concept of texture in image processing is discussed and the different analysis paradigms are analysed with reference to the type of texture they aim to model. The analysis of the large-scale texture of skin patterning forms the focus of the investigation; existing techniques are found to be inadequate for the description of this texture. A detailed investigation of the nature of skin patterning in lesion images is undertaken from which conclusions as to the requirements for modelling of this line segment based pattern are drawn. The abstracted representation is constructed in view of the need to capture the essential properties which would allow the measurement of disruption. A new technique is presented which is effective in extracting a representation of the quality of the skin line patterning.

6.1 Texture in Image Processing

The term *texture* in digital image processing generally refers to local spatial variations in image values, and aims to describe properties such as smoothness, coarseness and regularity. It is difficult to completely describe this notion either qualitatively or quantitatively [60, 78, 79]. Texture analysis can be divided into five methodologies as detailed below. The first three categories reflect early

(such techniques are still widely used and researched however, for example Liu *et al* 1996 [80] proposed a texture modelling strategy based on "periodicity directionality and randomness") research which sought to characterize the texture by quantification of 'properties' as described above. The fourth category aims to produce an analytical model of the texture and in the final category the texture is considered as a composite of primitive elements. In more detail the five categories are:

Statistical The texture is modeled as a statistical interaction of the values in a local area. Once a statistical texture description has been created a particular region can be tested for consistency with that texture by examining the difference between the area and the statistical expected values. Examples of this class of texture method are co-occurrence matrices and two-dimensional autocorrelation [81]. Statistical techniques have been extensively used in remote sensing applications and for land use classification [79, 81].

Spectral The texture is examined in the frequency or sequency domain (e.g. via the Fourier or Hadamard transform) [79]. In the transformed domain texture features are revealed, e.g. fine textures are characterised by high frequencies, coarse textures show as lower frequencies. In practice however, spectral methods are not very effective where the texture is not very regular in period or orientation [60]. Spectral techniques have been widely used for classification of remote sensed (e.g. LANDSAT) data [79].

Micro Structure The texture is defined by the response to a set of local area filter kernels. The kernels are convolved with the image, each serving to accentuate specific features of microstructure such as lines, spots and ridges. Parameterization of the texture is achieved by taking the windowed standard deviation over an area containing several cycles of the texture for each of the kernel convolution results. This scheme was developed by Laws [82], who proposed a set of nine 3×3 kernels but many other kernel sets have been suggested including the 3×3 Chebyschev and Sobel gradients [60]. The kernel convolution results can be analysed in terms of total texture energy in a local region as used by Laws, or as a texture themselves, for

example using a statistical analysis method, the co-occurrence properties of micro-structures could be analysed.

Model Based The observed texture is assumed to be the product of a certain set of parameters under an analytical model. The parameter set defines the properties of the observed texture. Parameter sets can be estimated from a given neighbourhood using the least squares method and then these are compared to known parameter sets for classification. Markov random fields have been studied extensively as such a model of texture, the discrete Gaussian version of the model being a simple linear combination of the grey levels in the neighbourhood plus additive noise [79, 81]. Linear autoregressive [79] and Fractal [81, 83] models also fit into this class, the former being useful where a causal neighbourhood for texture generation is required [16] and the latter where the texture has the property of self-similarity at different scales [81]. This form of texture analysis has been applied to a variety of natural scene segmentation tasks [74], however it is perhaps most useful where *generation* of artificial texture is required as in the generation of synthetic lesion images for verification of boundary identification performance [18].

Primitives and Building Rules The texture is assumed to consist of a number of primitive elements (sometimes called *texels* such as triangles and lines, and the spatial arrangement and orientation of these primitives forms the texture [79, 81]. The primitives are essentially arranged according to a set of rules. The rules can be expressed as a grammar over the alphabet of the primitive set, or can be described as a texture themselves, for example by using the co-occurrence matrices over the primitive set rather than pixel values, or by considering qualities such as periodicity and directionality. These techniques can be particularly useful in recovering 3D object shape from an image, the variations in size shape and density of the texels on the object in the image indicating its 3D shape [81].

Most existing techniques for texture analysis consider only grey-scale values however there have been some attempts to extend the notion of texture to use colour. Two distinct approaches have been described, the first simply applies standard

techniques to a number of colour bands in turn and combines the results [14] whereas more recently the interactions between colour planes have been included within the model [74]. Unfortunately the consideration of three planes (e.g. red, green and blue) vastly increases the complexity of the texture analysis, so only relatively simple methods and small neighbourhoods have been used.

6.2 Texture for Lesion Classification

The use of texture analysis for the analysis and classification of skin lesion images has not been widely investigated, however texture has been used in an attempt to improve segmentation results for skin lesions [14]. In this case co-occurrence matrices are used to evaluate the texture separately for two bands of a transformed colour-space which removes intensity for consideration by another method. Texture work has also been undertaken for the purpose of generation of synthetic lesion images for the evaluation of boundary finding techniques for skin lesions [18]. This work found linear autoregressive models to be the most effective texture method for the generation of skin and lesion texture which was visually similar to skin and lesion texture.

Much more work has been done in characterization of difference in textural roughness between the lesion and surrounding skin, where surface topography (profilometry) is measured in linear traces perpendicular to the first order skin furrows [20]. Profiles are often taken from imprint replicas of the skin surface using a mechanical profilometer which by necessity must move a stylus very slowly over the replica, an optical profilometer is under development which should greatly increase the speed of this process as well as allowing profiles to be taken direct from the patient's skin [21]. The texture of the surface topography can then be analysed using a number of standard methods such as kurtosis and fractal dimension in addition to measures such as peak count, profile depth and average peak distance and classification calculated from a combination of these measures.

None of the above techniques allows for even a partially complete description of the skin surface patterning in terms of skin lines. The image based texture tech-

niques described, concentrate on small neighbourhoods and consequently cannot model the skin lines as is well illustrated by their absence in the generated skin textures in [18]. Although surface roughness comparisons have proved effective in discriminating between benign naevi and melanomas [20] it is clear that this analysis is only considering a small part of the "skin line pattern" texture. Malignant lesions disrupt the surface pattern in more ways than simply altering the roughness or coarseness that has so far been investigated.

6.3 Skin Patterning Texture

The surface of most areas of skin (except palms and soles) is covered in a network of fine lines that are a product of the structure of the top layer of the epidermis. Clinical features that distinguish malignant melanoma from melanocytic naevus include disruption of the skin surface (erosion or crusting) and the presence of irregular clumps of abnormal cells in the upper dermis [1, 5]. These features can be seen in the disruption of the skin line pattern across the lesion, for example, the consensus statement of the USA National Institutes of Health [6] states that earliest melanoma can alter these skin markings.

It has been argued ([8]) that serious disruption of the skin patterning will only occur when there is significant disturbance of the structure of the dermis by a malignant lesion, however small scale disruption does seem to be detectable even for early malignant melanomas with little vertical invasion as seen in the results (section 7.5) of chapter 7.

6.3.1 Existing Techniques

Most existing techniques for texture analysis consider the intensity fluctuations in a neighbourhood as *texture*. Skin lines are macroscopic features composed of fine linear elements, this means that a relatively high resolution is needed for them to become visible. A neighbourhood for the detection of skin lines will then be relatively large, with the skin lines only featuring in a small portion of the

data. In addition, skin lines are not perfectly regular in orientation, spacing and thickness, and hence form only a pseudo-pattern in the neighbourhood.

Statistical and model based techniques will, by definition, consider sparse irregular features such as skin lines to be noise rather than part of the texture. The result of analysis by such methods is a model of the texture for unlined skin – as illustrated by synthetic image generation results [18]. Such techniques are most effective in characterizing the texture when the elements from which it is composed are small, otherwise it is necessary to extract the properties of the macroscopic texture elements by other methods [81].

Microstructure methods, for example, Laws [82], rely on the response to predefined kernels and suffer from their scale dependence, recognizing structures of fixed pixel size, and hence a change of scale or skin line thickness/separation will significantly alter the results. Similar issues also affect spectral methods. Under these methods skin lines could conceivably be examined by considering only a small frequency range relating to the line thickness. The position of peaks corresponding to these frequencies in the Fourier spectrum would then provide the orientation structure of the skin pattern. Spectral techniques are however inherently reliant on high regularity in the texture structure so that the differing thickness and separation of lines together with the considerable variation in orientation would make them unsuitable for this application.

Primitive and building rule texture analysis methods[1] are different to those described above as they do not explicitly examine a neighbourhood, however, the requirement for high resolution still applies as does the observation concerning the relatively small area influence of the skin patterning. Although the crossing of skin lines forms triangles and other polygons, these shapes cannot be easily described by a small set of primitives due to their inherent irregularity and consequently, primitive and building rule methods are difficult to apply to skin line detection.

Skin lines are poorly described by the normal definition of texture and poorly detected by the existing techniques. A different method is needed to analyse this form of texture.

[1] The skin pattern detection method proposed in this study most closely relates to this category in that line primitives are considered and their properties analysed.

6.3.2 Related Line Pattern Applications

The detection of patterns formed by the interaction of linear components is also important in mammography where malignant lesions are often characterised by architectural distortions such as radiating linear structures [84]. The line patterns found in these cases are similar to skin patterning in that they only exhibit limited forms of regularity, however the focused radiating patterns that are often at the heart of such detection techniques are not relevant to skin pattern indicators. The techniques that have been employed in this field are still of considerable interest where they are less specific in terms of the patterns they detect. A method based on multi-scale directional feature extraction followed by factor analysis has been shown to be effective in detection of stellate lesions whilst remaining uncommitted to a particular pattern [85]. At each scale the method assigns a single orientation to patches spaced evenly over the image as the orientation of the pixel in the locality of the grid point with the maximum response from the directional detector. Such an approach is not suitable for skin patterning analysis as more than one orientation is likely in skin patterning in each patch – this is discussed in more detail in section 6.3.3. The factor analysis approach would not be effective in detecting malignant skin lesions as the most common form of patterning disruption is randomization of orientations. Factor analysis could perhaps be used to generate a model for normal skin, however is should be noted that the patterning in skin has been found to be highly variable between examples and no simple common construct (equivalent to the radiating structure of a stellate lesion) has been found. Skin lesion analysis is better suited by methods that can be used to detect a disruption over the lesion area relative to the patterning found in the surrounding skin.

6.3.3 Profiling by Orientation of Linear features

Skin patterning is a result of the complex interaction of fine linear features and as noted before, the pattern is far from being regular in most senses. In order to detect when this pattern has been disrupted some characteristics of "normal" patterning must be identified. Although typical skin patterning has been de-

scribed as forming a rhomboidal pattern [5] visualization of this pattern requires only a subset of the lines to be considered, in fact in many cases it is perhaps more accurate to say that the pattern consists of rhomboids divided into two triangles (see figure 2.3). In addition the scale of the shapes seen varies considerably depending on body site, being much larger on the back of the hand than on the forearm, for example. A feature of skin patterning that does show consistency is the prevalence of a number of directions over an extended local area. These "preferred orientations" often extend over the entire skin area of a typical clinical image of a lesion. Where a lesion has disrupted the skin surface this consistency is lost, some lesion areas then have no features like skin lines but most showing a random arrangement of orientations. Skin line patterning is then a form of texture that needs a new detection and analysis method.

A measure of skin patterning disruption can then be conceived as the comparison of consistency of preferred orientations over the lesion area when compared to the surrounding skin. The comparison to the surrounding skin allows some tolerance to the many variables in image acquisition (such as scale, orientation, lighting and even skin type) assuming of course that the skin lines are still visible in the given situation! A static model of preferred orientation relationships in "normal skin" is certainly conceivable and, if nothing else could be used to provide a measure of confidence in any acquisition of skin patterning in skin regions.

The disruption measure indicated above can be calculated by forming an abstracted representation of the patterning properties of neighbourhoods spaced right across an image of a suspect lesion and comparing some sort of model for the properties of those found in the skin with those found in the lesion. This task can be divided into four sections:

enhancement by highlighting the pattern and removing as much and as many of the unwanted background features as possible.

profiling by reduction of a region to a description of preferred orientations.

consistency analysis by comparing profiles in a locality and either forming classes of similar profiles or characterizing that locality by self consistency.

evaluation by comparing the characteristics that represent the lesion area with those of the surrounding skin.

The first two are discussed in the remainder of this chapter and fall naturally under the heading of profile acquisition, the other two are the subject of chapter 7.

6.3.4 Detecting Skin Line Patterning: Overview

The first stage in the design of an effective detector for a feature is an understanding of the nature of that feature; what properties can be attributed to the feature and thus used in its detection? The skin lines that form the patterning to be detected have very little to offer in this respect; they have an inherent variable thickness and spacing which is accentuated by the requirement for some scale tolerance, a variable intensity which is only a slight deviation from the background and which is again affected by a tolerance requirement to variable lighting, and a variable orientation which needs to be detected.

The process described here addresses this lack of recognizable properties by reducing the "model" used to the simplest form, where a linear feature is simply a consistency of value at an orientation and highlights the skin lines by their small negative deviation from the local background intensity in a grey-scale image.

6.4 Enhancing Skin Patterning

In an unmodified image the skin patterning "texture" is only faintly visible as fine lines of slightly lower intensity than the skin or lesion they are passing through. The first stage in analysing this texture is to highlight this skin patterning and to remove the variation in local mean caused by features such as a dark lesion. The enhancement process was addressed in two parts, firstly the extraction of the raw skin line features was investigated, and secondly possible methods for improving the quality of the revealed pattern were considered.

6.4.1 Exposing The Raw Pattern

The skin line pattern was exposed by removing features which are the wrong scale to be skin lines. The two most effective methods found to perform this processing are as follows:

Smooth Model Subtraction The essence of this method is the creation of a template for gross feature removal. The template is created from the image by convolving with a 9x9 window mean smoothing kernel and this template is then subtracted from the original. The result (similar to high pass filtered version of the original) is then enhanced by histogram equalization and the values inverted so that the skin lines are seen as high intensities. The method is simple, quick and provides an adequate depiction of the skin lines network.

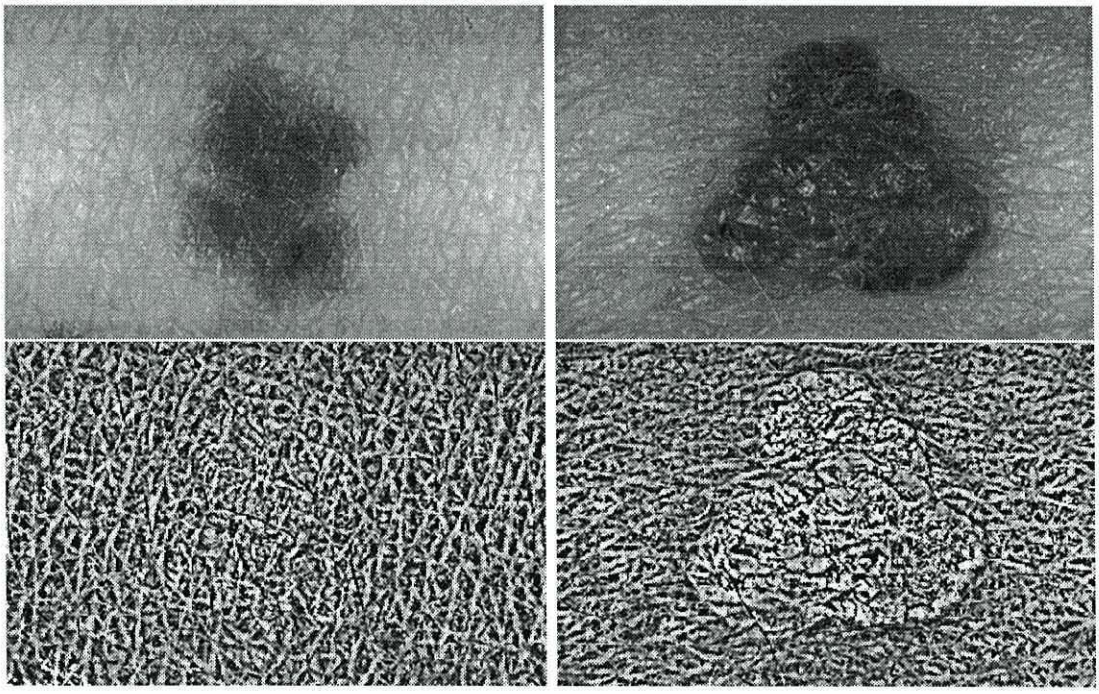


Figure 6.1: Skin line highlighting processing result for benign (left) and malignant (right) example lesions.

Figure 6.1 shows two examples of the preprocessing result. Unfortunately this step is sensitive to loss of contrast between the skin lines and their neighbourhood. Loss of focus causes loss of contrast and fine detail as discussed in section 4.4.2 and hence curvature in the subject, such as on an arm or ankle can affect the quality of the enhancement.

FFT based band-pass filtering Conceptually this is also a simple method as it simply involves identifying a range of frequencies that describe the skin lines and then using band-pass filtering to extract them from the image. Figure 6.2 gives an indication of the structure of a typical image by frequency. The application of low and high cutoffs of 0.22 and 0.55 respectively isolate the skin lines to provide a similar enhancement quality to that obtained using the smooth model subtraction technique.

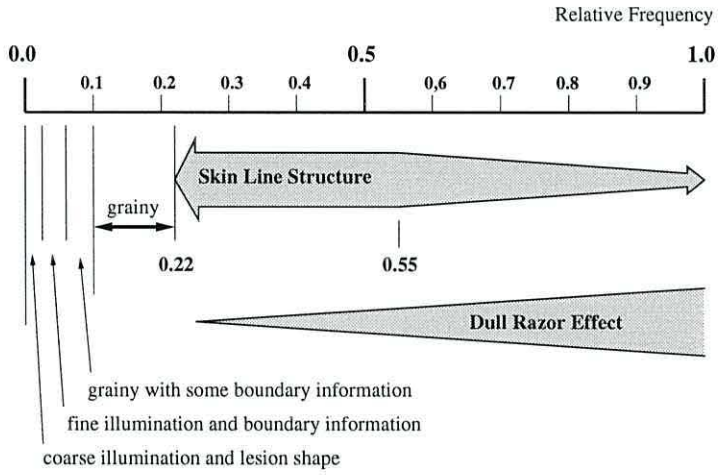


Figure 6.2: Structure by frequency for FFT filtering (0=frequency of zero (constant) 1=Nyquist frequency, ≈0.0365mm pixels)

This technique does not provide significant improvement over the smooth model subtraction and due to the computational complexity of the FFT transformations, this method was eventually discarded. The FFT transformation is in itself computationally expensive and furthermore requires that the region width and height are 2^n for some $n \in \mathcal{N}$. For the main image set used this implies a large amount of wasted calculation and a prohibitively long completion time.

The first method, smooth model subtraction, was selected and is used in the remainder of this study. However, as hardware performance improves, a reconsideration of the FFT method should be undertaken as it is perhaps more robust to, and amenable to adjustment for, image scale changes. It is also interesting to note that standard texture representations could also possibly be used to highlight the skin lines as a direct result of their inability to model them effectively. The deviation from the generated or expected value from the texture model ought perhaps to reveal skin patterning, however in practice this method

may be ineffective as the generated texture might not be well registered (i.e. synchronised) with that of the image. This method has not been investigated further as, aside from the synchronization issue, it would be far more computationally complex than the selected method which already produces seemingly adequate results.

6.4.2 Cleaning The Exposed Pattern

The raw skin line structure, as revealed, is still noisy and hence further cleaning of the result with the aim of producing a binary skeleton for the skin line pattern was addressed (figure 6.3). Skin patterning is often composed of more than one 'level' of structure, with more pronounced skin lines forming a pattern of rhomboids and half-rhomboid triangles, which are then further divided by finer lines into smaller triangles and other shapes. In many cases the most pronounced *primary* skin lines clearly follow only one approximate orientation, and the rhomboid-triangle shapes are seen only by also considering a secondary level of skin lines. The patterns formed by even finer lines are complex and preferred orientations become less consistent and harder to identify. In the ideal case extraction of only the more consistent larger scale (primary and secondary line) patterning is desired but such separation is difficult due to the natural variability of the skin surface. The lines from which the pattern is composed often have small breaks and are sometimes much less distinct than their contemporaries, and thus of comparable strength to those of the finer structures.

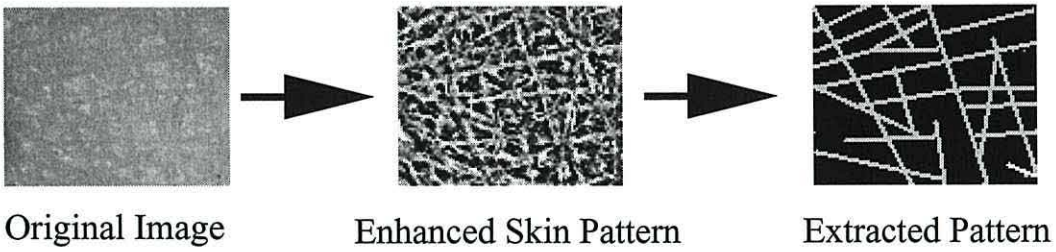


Figure 6.3: Illustration of skin patterning extraction concepts

Thresholding is simple to implement, but is not effective as it only accentuates the problems of faint components and breaks. The formation of the binary image

effectively makes small breaks more distinct and involves promotion or removal of the faint components. Using a global threshold, this process obviously cannot be performed in a manner consistent with the patterning level to which the faint components essentially belong, and a local approach cannot provide the solution as it is an understanding of the overall skin pattern and not local image properties which defines the correct action for a given line element.

Standard morphological routines for skeletonization were applied but perhaps unsurprisingly resulted in either an over complex pattern (due to the raising of all faint lines to the same importance as the main skin pattern) or a broken pattern (due to the small breaks and the removal of faint components). The non-uniformity of the line spacing is perhaps the most significant cause of this behaviour.

A method which traces ridge maximum values at the pixel level horizontally, vertically and at the two diagonal directions was devised. The final results, although showing some promise, shared many of the drawbacks of the morphological methods; especially in terms of the conflict between detecting faint components and ignoring the finer structure. Small breaks in the lines could be bridged by searching ahead a small number of pixels for a continuing line, however, this had a tendency to also bridge between many short curving line segments and form a false straight line from them.

None of the methods investigated were sufficiently robust and faithful to the skin patterning over a range of sample images to be used as part of the final process, and instead the results of the smooth model subtraction highlighting method were used directly. However, the clarity of the resulting pattern obtained where cleaning had been successful clearly indicated the potential of such processing and the value of future research in this area.

6.5 Regional Profiling for Skin Patterning

Once the skin line patterning has been highlighted and much of the background feature content has been suppressed the next stage in profile acquisition is then

to reduce the skin line patterning to an abstract representation that embodies the required property of local "preferred orientations" which can then be processed to produce a disruption measure.

6.5.1 Patch Based Processing

In the regional profiling process, small square patches of the enhanced image are processed separately. The patches are centred at points spaced evenly across the entire image and cover partially overlapping areas.

The identification of orientation in the patches, for the simple type of linear structure previously identified as the "model" for lines in the skin patterning, can be approached in two distinct ways: Two dimensional processing followed by analysis of the result over a range of angles or formation of a rotated version of the patch for each of the range of angles followed by single orientation analysis. Both methods will achieve the same results given that the understanding of sampling at an angle is consistent. The latter method was chosen as it is conceptually simpler, easier to verify that the process is functioning correctly and the construction of linear detectors is simpler. It should also be noted that this method has obvious potential for parallel calculation, with a separate process creating and analysing the rotated image for each angle in the range.

The patches are spaced finely enough that the data obtained reflects most of the variation in the image and yet coarse enough to avoid redundancy (measuring of the same information) and to keep the volume of data to a realistic size. The size of a patch is affected by similar criteria in that it must be large enough to include identifiable lines of the major skin pattern and not to be too sensitive to local faint lines and anomalies yet not so large that the profiles for adjacent areas are unlikely to differ significantly. Suitable values for the spacing and size are dependent on the capture resolution, however the effect of the variable scale seen in skin patterning should also be considered and it was found that in most cases skin patterning was best described by an area with side between 0.5 and 0.9mm. The majority of the testing was performed on images covering a 12.7mm $\times$ 8.5mm area digitized using 350$\times$230 pixels of side approximately 0.0365mm.

Testing with patches ranging from 0.5mm to 0.9mm showed little qualitative difference and therefore patches of side 19 pixels corresponding to approximately 0.7mm were used in all of the results presented.

The choice of angular sampling frequency is influenced by the need both to respond to the finest skin lines and to limit the number of the computationally expensive rotation operations which have to be performed and the size of the profile data for a whole image. The frequency needs to be sufficiently high to ensure that the skin lines will be close horizontal at some point in the rotation sequence. This need can be addressed by choosing an angle increment which results in a movement equivalent to a single pixel at the edge of the patch. For patches of side 19 pixels then, taking $1 = r\theta \Rightarrow \theta = \frac{1}{9.5} radians \approx 6°$. An increment of 5° was actually used in all the results in this study.

The underlying method used for the linear profiling of a patch is then as follows: The presence of skin lines at angle θ in the patch is evaluated by assessing the *line strength* in the patch viewed as a set of lines of data at this angle (figure 6.4). The term line strength is used here to reflect the level of evidence to suggests the presence of skin lines crossing the patch at this angle, and for the enhanced images this means the presence of a line of high intensity pixels. The result for a patch will be termed a *response profile* being a profile by angle of the level of evidence, given by the response of an estimator, for skin lines at that angle.

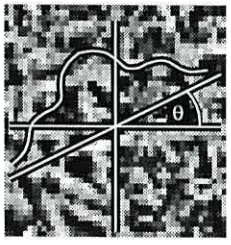


Figure 6.4: Assessing line-strength at an angle (curve shows autocorrelation of data – line strength by self-similarity)

This assessment is achieved by the following method: let I be a square patch of side d in the enhanced image. A slightly larger [2] area R (side $\sqrt{2}d$) centred around I is then rotated by θ, about its centre. An area I_θ the size of I is then

[2] The rotated area is expanded to ensure that there will always be enough data to fill the extracted area I_θ

extracted from the rotated image R, and the rows of I_θ are assessed for similarity.

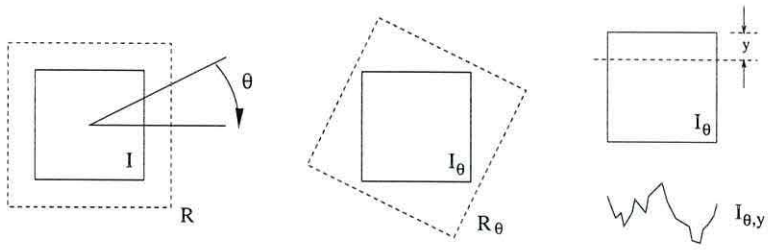


Figure 6.5: Method for re–sampling the image patch at an angle

6.5.2 Line Strength Estimators

Two estimators have shown promise in the assessment of line strength in the rotated patches: the first uses the 1D autocorrelation and the second looks directly for consistent values in the rows of the rotated image. For the purpose of the discussion of these methods it is worth noting that the requirements for selecting a patch size together with the irregularity and sparseness of skin lines means that it is unlikely that patches will contain many skin lines and it is unlikely that these lines will exactly match in orientation.

Autocorrelation method The autocorrelation function can be viewed as measuring the degree of similarity of the data when compared to a copy of itself shifted over a range of displacements. The autocorrelation function $\phi_s(k)$ is given by

$$\phi_s(k) = \frac{1}{p_s} \sum_{i=k}^{N-1} s(i-k)s(i) \qquad \text{where } p_s = \sum_{i=0}^{N-1} s(i)^2 \text{ and } s(x) : 0 \leq x < N-1 \tag{6.1}$$

For a given signal s, the "width", w of the function ϕ_s is a measure of the self-similarity of s. Thus $w(\phi_{I_{\theta,y}})$ is a measure of the similarity of row y in I re–sampled at angle θ.

The width of ϕ_s is currently evaluated as the point where the function first crosses 0.6 (issues concerning width evaluation and the choice of this threshold are discussed later in this section). Since ϕ_s is a discrete function of only a small number

of points, simply defining the width as the smallest k such that $\phi_s(k) < 0.6$ will not allow for much variation. For that reason linear interpolation is used between k and $k-1$ and is given by,

$$w(\phi(s)) = \frac{1}{d}\left((k-1) + \frac{\phi(k-1) - \phi(k)}{\phi(k-1) - 0.6}\right).$$

A single patch at a single angle therefore results in a set of d autocorrelation widths. These values are sorted into decreasing order and the lower half (containing the results for the 'noise' between the sparse skin lines) is discarded. The median of the remaining data is then taken to represent the strength of skin lines in the patch at the angle. The output is a vector of these strengths which forms the profile for skin line strength versus angle.

Figure 6.6 shows response graphs for sample patches from one of the malignant melanoma trial images (the MIS as shown in figure 2.5). A profile for some neighbouring patches is shown in each case. The first graph shows data from a normal area of lined skin with only one major preferred orientation close to horizontal (seen as a peak in response at either end of the plot). These skin area responses show the marked similarity that indicates that the patterning has not been disrupted. The second graph illustrates the behaviour found in malignant lesion areas, the three responses shown differ in character showing random 'noise' patterning.

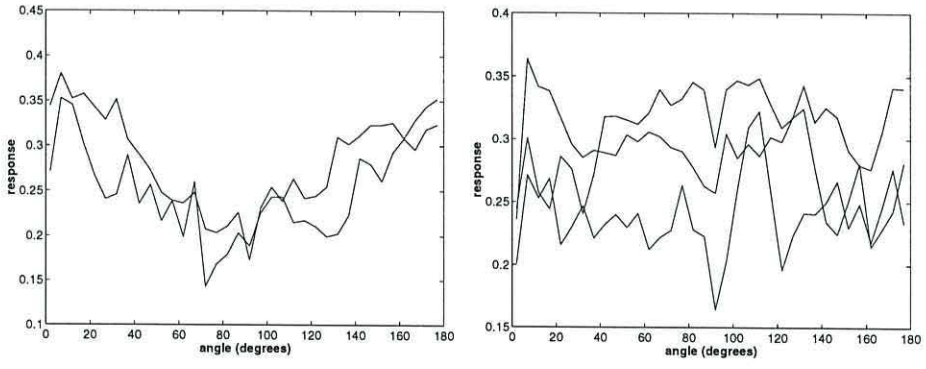


Figure 6.6: Sample responses for autocorrelation method: patterned skin (left) and malignant lesion (right).

Evaluation of the self-similarity expressed by the autocorrelation of the oriented rows is a complex issue as it is not easy to construct a width measure which will

extract the information relating only to our desired concept of similarity. The general form for output from the autocorrelation of a single row "signal" can be viewed as a decreasing graph starting at 1 and then decreasing rapidly at first and eventually reaching (almost) zero. High self-similarity manifests itself as less rapid decrease with a row of identical values producing a linear result (figure 6.7).

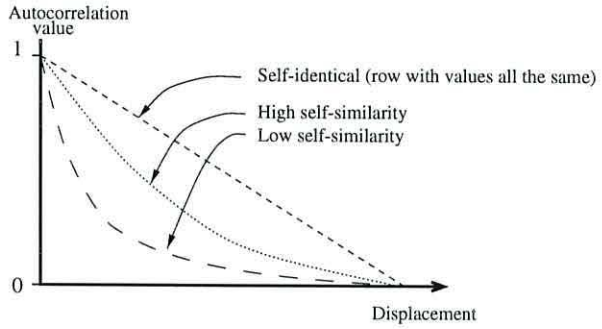


Figure 6.7: Basic autocorrelation function properties.

Unfortunately (for this application) this is not the whole story, autocorrelation responds to any form of periodic self-similarity and in such a way that the function does not smoothly decrease and can even increase. Figure 6.8 shows the autocorrelation result for two artificial rows containing periodic data together with the linear result of uniform data. The maximum value for discrete peak data is also shown. The metric employed to quantify the width w was chosen in an attempt to minimize the effect of this response to periodicity. Many standard width metrics such as RMS width are seriously affected by the periodic surges as all the data is used, including that for higher displacements where the surges become more dominant. The threshold-based metric does not suffer as greatly from the periodicity providing the threshold chosen is sufficiently high that the autocorrelation value meets the threshold before the surges are significant. However, the threshold must be low enough that there is detectable variation in the width value it provides given the normal situation of high but decreasing initial slope. The value of 0.6 seems to provide the best balance between these requirements (see figure 6.8).

The method as a whole performs quite well in terms of profiling line structure, producing a good response to the type of similarity we wish to detect, however it suffers from two major categories of drawback, undesirable high response in low contrast patches and "false" response related to the way autocorrelation responds

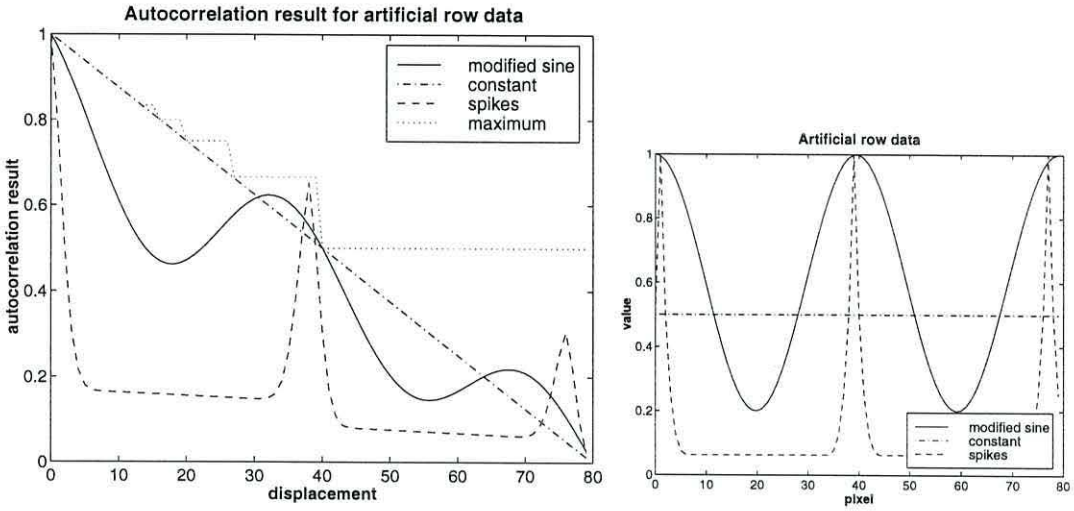


Figure 6.8: Periodic properties of autocorrelation.

to periodic similarity.

Low contrast patches result in a high response at all angles, simply because such patches are indeed self-similar at all angles. This is most undesirable in such areas of the enhanced image that are dark as detection of skin lines there would be misleading. Low contrast in a patch is simple to detect so this problem could be dealt with by marking such data as suspect or invalid. A related problem which again has a simple solution is the fact that the autocorrelation of a row of constant zeros is constant zero rather than the linear ramp of any other constant, again this is easily dealt with by manipulating the input data range or by masking, but nonetheless it demands mention.

The second major category relates to the effects of the periodicity response. Although the threshold method does greatly improve matters there are real-image situations in which "false" response attributable to periodicity does occur. A good example of such a situation is where skin lines with only one significant orientation and similar spacing are rotated to a vertical orientation. When the horizontal autocorrelation is evaluated, this results in rows with a wave-like pattern of consistent frequency and therefore a periodic surge in the autocorrelation response (similar to, although less pronounced than, that of the modified sine in figure 6.8) and hence the likelihood of a large width result and false high response recorded at that orientation. This effect can be clearly seen for the example in

figure 6.10 which shows response corresponding to the lines in the patch at $\approx$ 47° together with significant response at $\approx$ 47+90° – perpendicular to the lines. Furthermore, false response can occur at orientations cutting across two or more lines where the wave-like pattern is formed as an oriented row encounters each line in turn, thus a patch containing three lines can have two such "harmonic" responses either side of the true response. These responses are not so dependent on regular spacing of the multiple lines as the false perpendicular range is, and so would be a greater problem were they not at a much lower level since only a limited number of rows are affected. These "harmonic" responses may explain the "shoulders" seen either side of the true response in the figure.

Consistent Value method The consistent value approach looks directly at the values that make up the rows of the rotated patches. The investigation of the nature of skin patterning and the lines from which it is composed suggested that a detector should respond to consistent values of pixels rather than any more detailed model of a line and it is this that prompted the experimentation with the consistent value method. Consistency of value implies a low range and so $1 - range(I_{\theta,y})$ (where $range(l)$ is the difference between the maximum and minimum values in the data line l normalized to lie between 0 and 1) can be considered a measure of consistency.

This produced results that were surprisingly similar to the autocorrelation method but highlighted the problem with detection of smooth dark areas as skin lines; Low intensity smooth regions tend to have less variation than high intensity smooth regions and therefore result in a lower range and a higher response, the effect of which is the apparent detection of strong skin lines at most orientations in a dark smooth region.

A re-examination of the aims for the detector provides a simple but effective solution to these problems; The skin line highlighting stage aims to produce *bright* lines to represent skin patterning and so we should restrict our detector to bright lines with little variation. This can be achieved by simply taking a minimum value for a particular row as its line strength - when this is high, the entire row must be a bright line and when it is low, there is no line or the line is broken at some point. This method will be referred to as Consistent High-

value Profiling (CHP). Small breaks in the lines must be allowed for and this is achieved by taking the mean of the lowest three values as the line strength. Several other break compensation ideas were tried including simply taking the lowest-but-one value, taking the mean of more values and taking the Gaussian-weighted mean of the lower values, the first was not as effective in compensating for breaks, whereas the second over-compensated and began to detect portions of two separate lines as a whole, Gaussian weighting gave no noticeable advantage and involved considerable computation overhead.

As with the autocorrelation method, a single patch will result in a set of d line strength evaluations - one for each row of the rotated patch. Again these values are sorted into decreasing order, but this time a much larger portion of the lower values are discarded. This is because the detector now only responds to consistency of high value and not consistency at any value. The response for the patch is calculated as the mean of the highest $\frac{1}{5}$ of the row responses, the rationale for this figure being that only a small number (one or two) of skin lines are expected in a patch and these are usually less than three pixels in width, therefore a patch of side 19 pixels containing two lines should be expected to produce $2 \times 2 = 4$ rows of bright pixels at the orientation, $\approx \frac{1}{5}$ of the rows. The mean value is used rather than the median as it is computationally less expensive and produces comparable results (since outliers are rare with this strength measure).

Figure 6.9 shows response graphs for the CHP method for the same melanoma image used for the autocorrelation examples in figure 6.6. Again a profile for some neighbouring patches is shown in each case. The first graph shows data from a normal area of lined skin, the responses show the marked similarity that indicates that the patterning has not been disrupted. The second graph illustrates the behaviour found in malignant lesion areas, the three responses shown differ in character showing random 'noise' patterning.

Figure 6.10 shows a comparison, for a patch in the skin area of a different example (from the top left quadrant of the melanoma example in the top left of figure 4.5) of the profiles obtained through the CHP method and the autocorrelation method skin area in another melanoma example. The CHP method provides a profile clearly reflecting the major line orientation structure expected from a

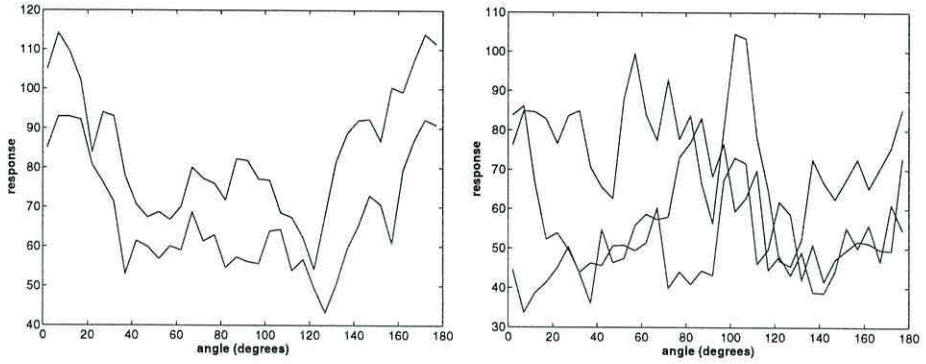


Figure 6.9: Sample responses for CHP method: patterned skin (left) and malignant lesion (right).

visual inspection of the pattern enhancement stage and of much increased quality when compared to the autocorrelation method. The orientations are more reliably detected and there is less apparent noise and spurious detection.

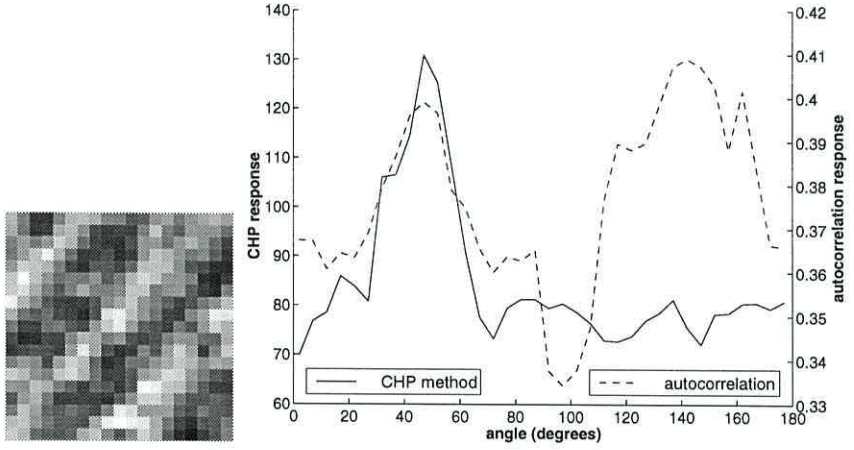


Figure 6.10: Response comparison for a given patch of patterned skin: The patch analysed and the profiles from CHP and autocorrelation methods.

6.5.3 Demonstration of Profiling

Once a profile has been constructed for patches spaced evenly over a whole image, the profiles can be reconstructed into an image showing the characteristics that were detected[3]. This provides an effective demonstration of the validity of the

[3] For the pictures shown here the profiles for the individual patches are individually contrast stretched, this improves the visual quality but in cases where no linear feature is found (rare)

process when applied to artificial and real images (figure 6.11).

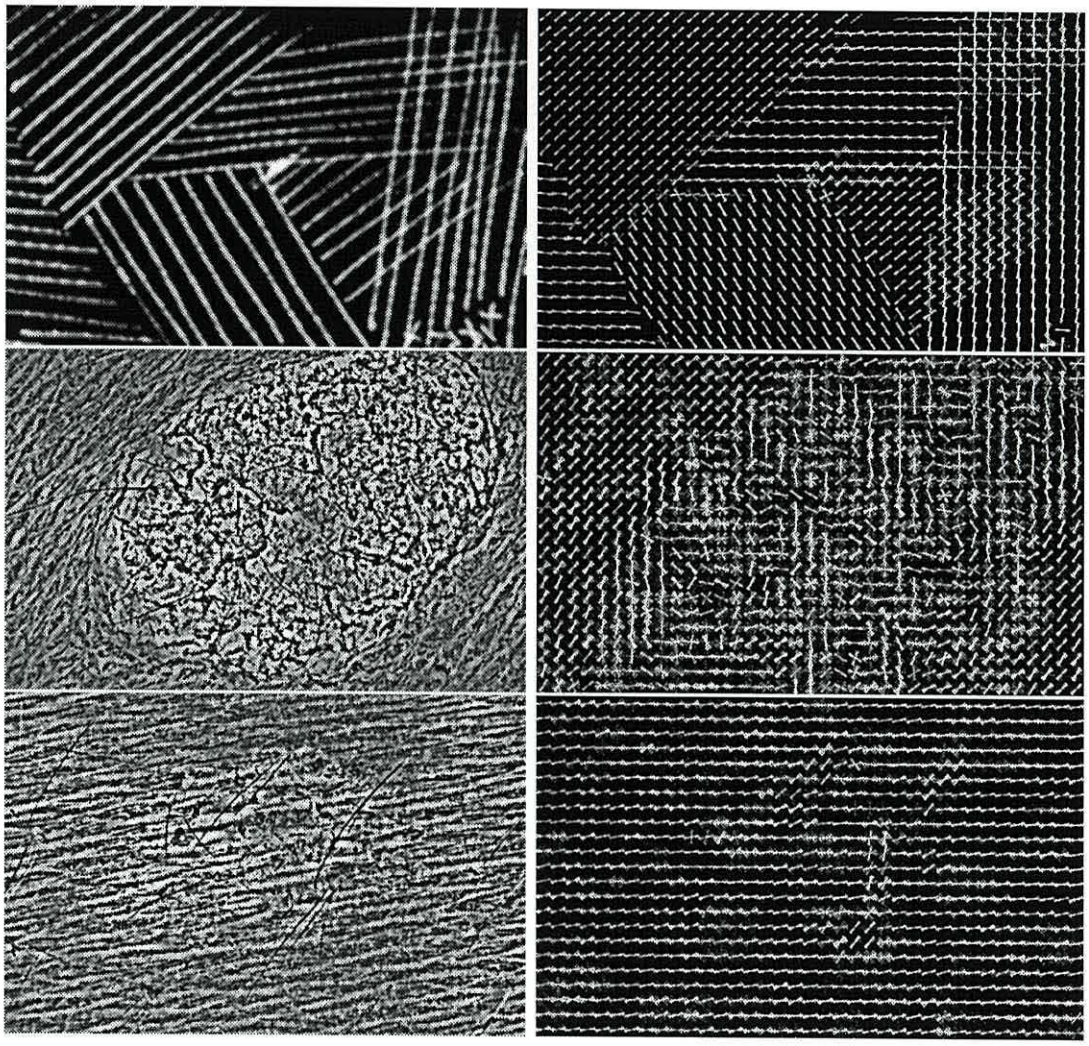


Figure 6.11: Profile representation: original line image and reconstructed profile image for: top, an artificial image, middle, a malignant melanoma and bottom a benign compound naevus. (The original lesion images are shown in the top left of figure 4.5 and figure 7.18 respectively.)

The artificial lines image is mapped successfully even where more than one direction of line is present in the source image and when the reconstruction does not use local enhancement a clear relationship between line strength and profile strength is seen. In the real image cases the skin line pattern profile is also extracted successfully – clearly seen as the diagonal lines for the skin surrounding the melanoma and the near horizontal lines over the majority of the benign case.

can lead to presentation of minute features and noise as true linear features.

6.6 Conclusions

The pattern of fine lines criss-crossing the surface of normal skin forms a macro-scale texture. Changes in this skin patterning texture have been identified as indicative of early melanoma. A measure of disruption of skin line patterning would then clearly be of benefit for a computer based diagnosis support system. Existing research relating to skin patterning texture extends only to changes in properties of skin surface profiles taken perpendicular to the direction of the main skin lines. It is clear however that considerable potential for diagnostic information exists in the disruption of the directional pattern formed by the skin lines.

The nature of skin patterning means that it is poorly described by existing texture techniques and consequently a new skin patterning detection method has been developed. An abstracted representation is proposed for this patterning and the linear elements from which it is composed. This model aims to capture the essential properties of the pattern and highlight the loss of order seen over disruptive lesions.

The new detection method forms a matrix of profiles (by angle) of 'line strength', in effect a map for the preferred orientations of the patterning on the image. The process begins with a simple method to highlight the skin lines in the lesion images by removing large scale features such as a dark lesion. Patches spaced evenly over the entire image are then analysed and each produces a profile representing the character of the local skin patterning. This process has been shown to be effective in relation to test patterns and in reflecting the observed skin patterning in real lesion images.

The next chapter addresses the analysis of skin pattern profile map data to provide the quantitative feature data needed for the automatic diagnosis support system.

Chapter 7

Texture – Skin Pattern Analysis & Disruption Evaluation

This chapter addresses feature analysis and evaluation using the skin pattern representation extracted by the method in chapter 6. Quantitative interpretation of the patterning information is vital if it is to be useful in an automated detection system. The provision of a metric for the quantification of disruption is therefore considered. Preliminary results for a number of analysis techniques prompt a number of changes and enhancements to both the extraction, analysis and evaluation methods. The final results show the effectiveness of this texture analysis and disruption feature evaluation in relation to both visual assessment and diagnostic performance.

7.1 Analysing Skin Pattern Profile Results: Initial work

The methods described in the previous section produce a vector for each area of the image which represents the characteristics of the skin lines found there ([86]). A visual representation of the skin patterning data acquired through the method can be simply generated (figure 6.11). This pictorial representation shows

the regularity expected in normal skin areas, its continuation over most benign lesions and its marked disruption over malignant lesions. However, automated analysis and quantification of the disruption is not simple — the matrix of profiles is difficult to interpret and an analysis method is required [87]. An indication of the potential for automatic evaluation of pattern disruption was required and this was pursued utilizing the earlier *autocorrelation* based profile technique (all results later in this chapter use CHP data). Four classification methods were investigated at this stage, two neural networks, a region agglomerative classifier and a variance based measure. The two neural networks, the Self Organizing Feature Map and ART-2a, were chosen as well known examples of unsupervised, self organizing classifier networks. This type of network was required as the profile properties considered as characteristic of 'normal' skin patterning relate to similarity between neighbouring profiles rather than to fixed features of individual profiles, the specific networks were chosen as examples of pre-defined and dynamically created output class configurations respectively. Each of the four methods is composed of two distinct stages: the analysis/classification stage where the profiles are compared locally or grouped by similarity and the evaluation stage where this data is converted into a meaningful measure of disruption or non-conformity between skin and lesion area profiles.

7.1.1 Local Variance

The first method uses the local variance in each plane as a measure of the neighbourhood similarity in response for a given angle. The mean of these variance results for a given patch provides the metric for an overall similarity measure. The 3x3 windowed variance is calculated for each plane (all profile values at a particular angle) and the mean of all the variance results for a given patch provides the metric for the similarity of responses in the neighbourhood of that patch.

7.1.2 Adaptive Resonance Theory

The ART networks [88] have also been considered as a method for categorizing the output from the skin line detector. In particular the ART-2a version is used.

It performs unsupervised on-line categorization of input patterns into a number of classes.

It is a two layer network in which the input layer performs normalization and noise reduction on the input data. The two layers are connected by two sets of weights which fully connect each input node to each output node. The normalized input data is passed to the output nodes via one set of weights and the output nodes "compete" with each other to determine the one which best represents the input. All except the winning node are then "switched off" and hypothesis testing is performed within the output layer to ensure that the chosen output node sufficiently well represents the input (the other set of weights are used for this). If the hypothesis test fails then the chosen output node is switched off and the competition repeated without this previously chosen node. This process is repeated until either a matching node is found or all available nodes have been tested. In either case a learning procedure is started. If a matching output node is found then the weights between that node and the input nodes are adjusted to make the node more closely represent the input. If no sufficiently well matched output node is found a new node is recruited and its weights are made to represent the input data. This method uses the cosine of the angle between vectors as the measure of similarity and as stated has built in normalization prior to comparison.

7.1.3 Self Organizing Feature Map

The Self Organizing Feature Map (SOFM) [89] is a single layer neural network in which each output node is connected to all of the inputs. It makes use of competitive learning to produce a single activated node in the output layer for a given input vector. The effect is to cluster similar input vectors together, with similar numbers of vectors being "attributed" to each of the output nodes.

The process is iterative; classification begins with the production of a global ordering of the data and then processing becomes progressively more localised as the iteration proceeds. This is achieved through a shrinking "neighbourhood" of network weight alterations in the learning process. The output class for each output node is taken to be the set of input samples which are best represented

by the weights associated with that node.

In this application a 1-D SOFM with Euclidean distance as the similarity measure has been used. The SOFM requires the number of output nodes to be fixed. This number of output classes must be chosen so as to sufficiently, but not overly, divide the number of profiles in a matrix for an image (as discussed further in section 7.2), and this was set at 32 for all the lesion image processing and results presented.

7.1.4 Multi–stage Region Agglomerative Clustering (mRAC)

The final classification technique applied is the novel agglomerative clustering technique presented in chapter 5. Although the mRAC [77] process was originally developed for analysing colour images it can be used to cluster any image composed of vector data.

Clustering methods aim to partition data into groups consisting of elements which possess similarity. The mRAC process differs from the others presented here as it has two criteria for similarity; the first is explicit – the distance measure used in judging the similarity of two clusters, and the second is implicit – two clusters must be spatially adjacent to be considered uniform. This implicit similarity criterion has an important defining influence on the final clustering solution produced. In discussions concerning mRAC, clusters are called regions.

The uniformity criterion initially used for the profile data relates to the Euclidean distance as before. All parameters were kept constant for all the disruption evaluation results given in section 7.2. The exact parameter settings for profile data are not critical, different but similar parameter sets generate results differing in character but not in validity. The setting of t_{optm} obviously has the greatest effect on the final segmentation. The thresholds used here were; splitting $t_s \approx 80$, conservative merge $t_{cm} \approx 90$ and optimal merge $t_{optm} \approx 230$. These thresholds relate to the expected range of the uniformity criterion d which theoretically (for the 37-element profile data) had a range of 0 - $\sqrt{37\text{x}(255)^2}$ (≈ 1550). The

cleaning stage was employed for all regions smaller than 0.1% of the image area.

7.1.5 Demonstration of Classification Techniques

In order to illustrate the intended operation of the technique a test image of hand-drawn lines in various orientations was created. The line response profile was calculated and the analysis methods applied. Figure 7.1 shows the segmentations obtained superimposed over the test image.

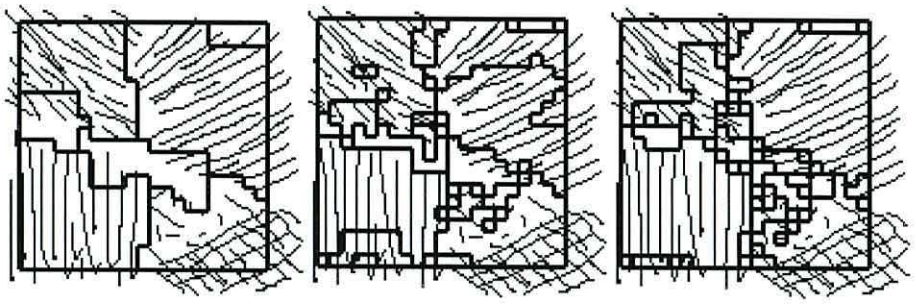


Figure 7.1: Segmentation of a test image containing had drawn lines at various orientations. Left: mRAC, centre: SOFM with 8 classes and right ART

Initial testing of the ART method on the artificial line data produced acceptable results apart from poor classification in areas where many consistent regions met (as seen in the centre of the image above). However, when ART was applied to the autocorrelation based profile data produced from real skin images it was found that the loss of the magnitude information during the normalization process made the resulting segmentation confused and difficult to interpret visually and lead to inconclusive results in automatic evaluation of pattern disruption. As a result ART was not considered further in this early evaluation phase. Investigation since then (and discussed later) has shown that ART is effective in the interpretation of the much "purer" profile data produced through the CHP (section 7.4.6) method especially where the internal normalization is controlled.

7.2 Evaluating the Classification Results: Initial Work

A quantitative measure of the actual disruption to the skin patterning caused by the lesion is required. This measure is based on the relationship between the classification results and a mask which defines the lesion and skin areas (obtained automatically via the intensity based edge-focusing segmentation algorithm where possible). In general, the essence of the disruption measure can be described as: where the texture based segmentation reveals the lesion then the skin structure has been changed by the lesion.

Each of the three successful profile analysis techniques (local variance, mRAC and SOFM) produce results with significantly different characteristics. As a result, each requires a different evaluation method. Figure 7.2 shows from left to right, the original image, mRAC, SOFM and local variance process results for two melanomas and two compound naevi from the test set. In general the points to note are:

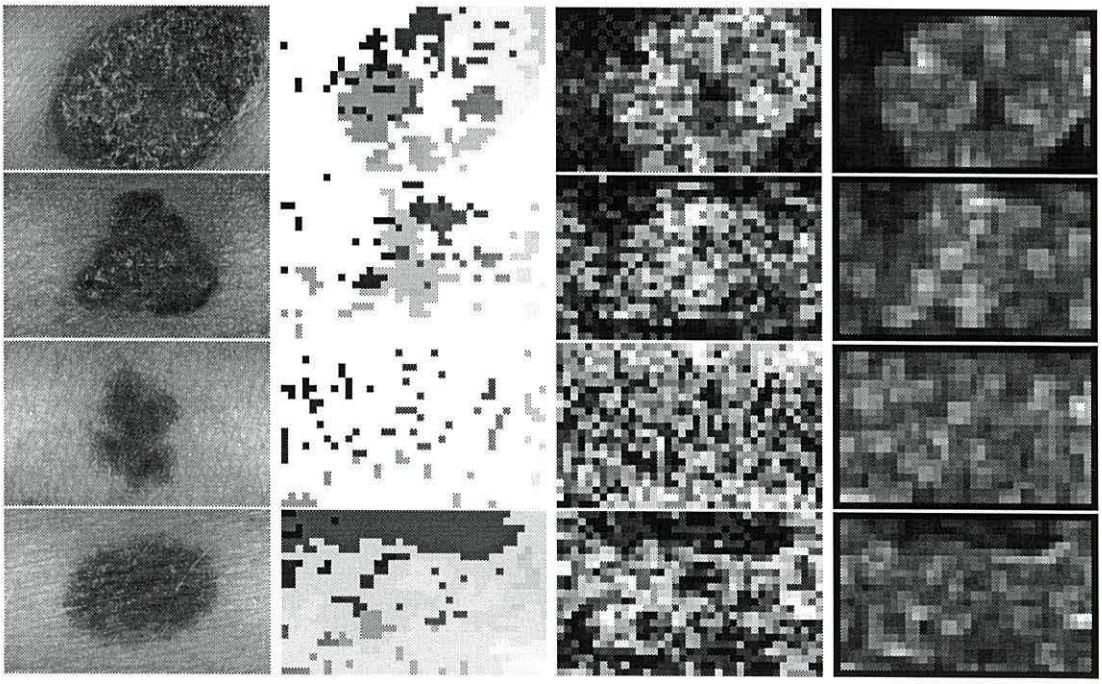


Figure 7.2: Top to bottom: two example melanomas, two example benign naevi, Left to right: original, mRAC, SOFM (32 classes), local variance.

mRAC. Malignant lesions appear as a group of regions that are distinct from those that make up the skin area. In benign cases much of the lesion (often 85% or more) is classed with the surrounding skin.

SOFM. The SOFM classifications are characterised by the skin and lesion areas being most distinct where the lesion is a melanoma. In some cases the skin lines are actually lost altogether over the lesion (rather than being disrupted).

Local variance. Malignant lesions are seen as areas of high variance in comparison to the surrounding skin area, indicating that the skin line patterning has been disrupted. In benign cases this contrast is not found.

7.2.1 mRAC evaluation

The evaluation of the mRAC classification needs to detect the presence of regions in the lesion area that are not skin classes. For this purpose the concept of a *true-skin* class is introduced and defined as any class that accounts for more than t_{ts}% of the skin area. The lesion disruption figure is then calculated as the area of the lesion that is not accounted for by a true-skin class, and expressed as a percentage of the lesion area. The setting of t_{ts} does not seem to be critical, the value 10% was used for the test set.

There are several situations that reduce the confidence that can be placed in this measure: if the area of either lesion or skin is insignificant in the image and if the true-skin classes do not describe a sufficient portion of the skin area. The latter situation did not arise in the image set used, 71.5% being the minimum, however in several images the lesion is relatively small.

Although it cannot be seen in the static results, it should be noted that for melanoma lesions the method produces regions corresponding to skin areas before (at a lower threshold) those corresponding to lesion areas. This may provide another method for obtaining the skin line disruption information, but has not been investigated in this study.

7.2.2 SOFM evaluation

The SOFM evaluation needs to detect where the lesion area uses classes distinct from those that make up the skin. This is different to the measure for mRAC because SOFM has to work with a fixed number of classes even where there is little variation in the data. The effect of this limitation is seen in benign cases where the skin and lesion (which are generally a single class under mRAC) are distributed randomly between the 32 classes. In addition, the limitation requires that the number of classes be set to sufficiently, but not overly, divide the number of samples.

The region numbers (seen as different intensities) assigned by the SOFM form a continuum and as such the fact that the lesion is bright (composed of regions with high numbers) and the skin is dark (low numbered regions) is meaningful. However, it is not clear exactly how the region number corresponds to the skin patterning and so for the current evaluation, any information in the relative numbering has been ignored.

The result aims to reflect the percentage of classes that have their only significant appearance in the lesion area. Where the lesion uses the same classes as the skin this will produce a low result, and where the classes used are distinct the result will be high. The actual figure computed once again uses the concept of true-skin, however, this time the result is given by the percentage of classes that are not true-skin. Since the number of classes is fixed, each class must account for the same portion of the entire image, therefore, all classes that do not represent a sufficient portion of the skin to be considered true-skin can be viewed as having their only significant appearance in the lesion.

Since each SOFM class will theoretically have the same number of members, it will cover the same area in the entire image. This means that the number of classes for SOFM can be chosen to produce a desired class coverage and the value of t_{ts} can be estimated to *exclude* all classes that are representing mostly lesion area. For the results below SOFM with 32 classes was used, this means a coverage of 3.1% of the entire image for each class, and assuming the lesion covers $\frac{1}{3}$ of the image, suggests a t_{ts} of 2% (actually 1.8 was used).

7.2.3 Local Variance Evaluation

The local variance needs only a simple evaluation. The mean value for the disruption (local variance) is calculated for both the skin (μ_S) and the lesion (μ_L) areas, and the ratio $\frac{\mu_L}{\mu_S}$ then represents the contrast between the areas. A high value would indicate high disruption caused by the lesion.

Due to the nature of the local variance operator, all the data points on the edge of the local variance image must be excluded.

7.3 Results and Discussion: Initial Work

The images used in the trials of the skin pattern techniques throughout this chapter come from the same set described in section 5.7 which contains examples of several types of lesion including malignant melanoma and compound naevi. The original images are 24-bit full colour digitized from slides with approximately 4 micron pixels. Each image is sub-sampled and converted to grey-level intensity to produce 350x230 source images.

Results are presented for ten images, five containing a malignant melanoma and five containing a compound naevus. The images were chosen to have a reasonable skin area (visually showing patterning) surrounding the lesions.

In order to properly evaluate the measures in relation to skin line patterning disruption, the results from each of the three evaluations are compared to a visual estimate of the disruption as well as to the histologically diagnosed type. The comparison to the visual estimate is needed for two reasons: firstly, the benign naevi in the test image set often have abnormal features (the reason for them having been referred and therefore being available in the image set), features that can include surface disruption and mean that the lesion perhaps *should* be mis-classified and secondly, the visual estimate comparison is vital in ensuring that the described measure does indeed quantify the skin patterning situation as it is visually perceived and not some other feature or combination. The scale used for this subjective estimate compares the patterning in the skin area to that

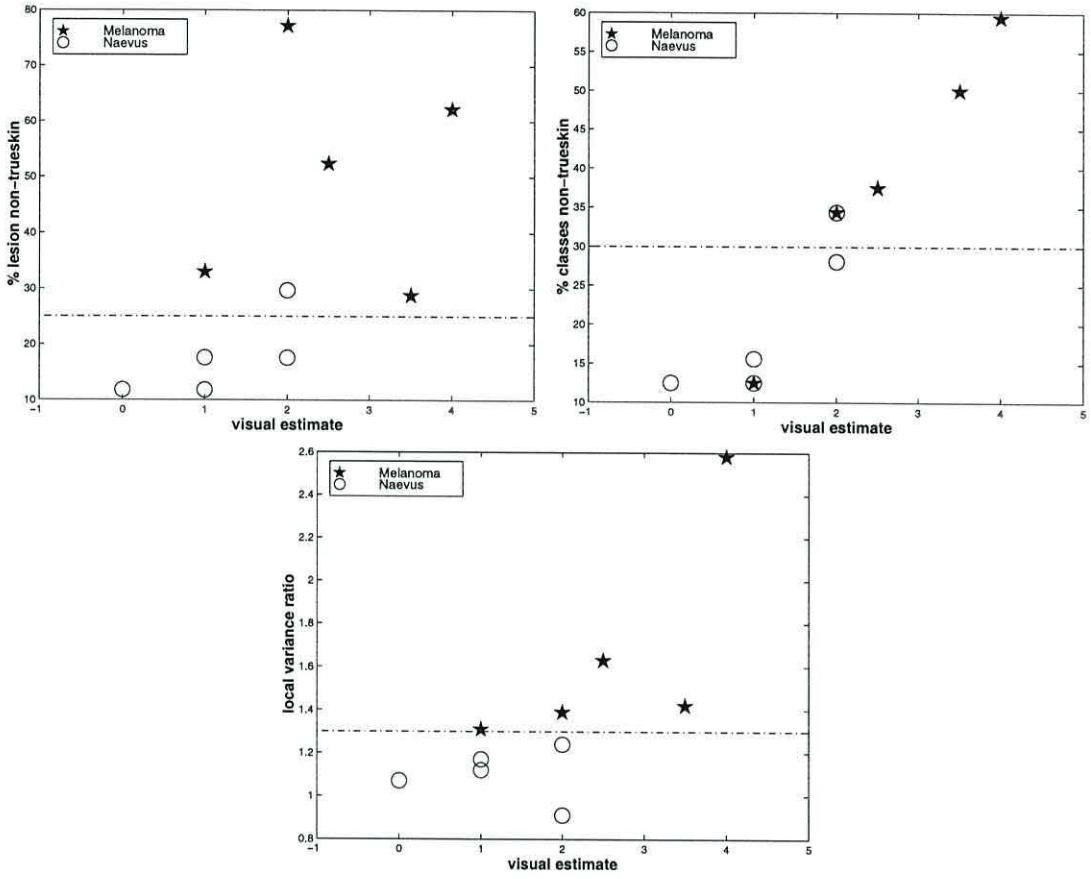


Figure 7.3: Evaluation results plotted as scatter against the visual estimate. In order; mRAC %lesion non-trueskin, SOFM %classes non-trueskin and local variance ratio of means.

over the lesion and relates to land-marks descriptions as follows:

disruption estimate	**description**
0	no noticeable patterning difference
1	slight pattern alteration apparent
2	notable change in pattern qualities
3	severe pattern disruption
4	complete loss of any pattern regularity over the lesion

Multiple observers were used and the mean of the scores (which were actually quite consistent) was taken as the final value.

Figure 7.3 shows the result for mRAC, SOFM and local variance evaluations scattered against the visual estimate. The correlation between the measures

and the visual estimates are generally good, and particularly so for the SOFM evaluation. The separation of lesion types is always at least as successful as the visual estimate, and in the case of the variance evaluation, the separation is perfect for the image set used.

In all, these results show that skin pattern profiling can provide information through local variability analysis or classification which can be reduced to a meaningful measure of skin patterning disruption. Each of the three measures reflects visual estimation of disruption and achieves a good separation of malignant and benign lesion types [90].

7.4 Developing Profile Analysis and Evaluation

The initial results demonstrated a promising correspondence between the instinctive visual appraisal of skin pattern consistency and disruption and the applicability of the analysis-evaluation combinations as a diagnostic indicator. However, many issues and possible avenues for improvement were highlighted and an investigation of these together with the integration of the CHP method for profiling are the subject of this section.

7.4.1 Profile Enhancement and Noise Reduction

Prior to analysis of the matrix of profiles the possibilities of noise reduction and signal enhancement need to be addressed. Although the raw CHP data is effective in its representation of the observed skin patterning character it would nonetheless improve the quality of classifications and local difference analyses if this underlying character as reflected in the profile could be enhanced and background noise could be reduced.

As previously noted, skin line patterning is a multi-level pattern in that some of the lines from which it is composed are more prominent than others. The preferred orientations result from the combination of lines of similar prominence

(although sometimes a much fainter line is needed to complete the pattern in a given area). However, the absolute strength of lines in the enhanced images is not always constant between different regions due to imaging effects such as loss of contrast in the enhanced image resulting from loss of focus in the original for example. The patterning is usually still distinguishable in the sense that the orientation of the most prominent "primary" lines in one region reflect those of other regions even where the relative strengths of these skin lines are considerably different in the enhanced image and the relative strengths are also preserved as in figure 7.4 where the skin patterning of strong lines (near vertical) crossed by slightly fainter lines (near horizontal) remains in proportion regardless of the overall intensity of the region.

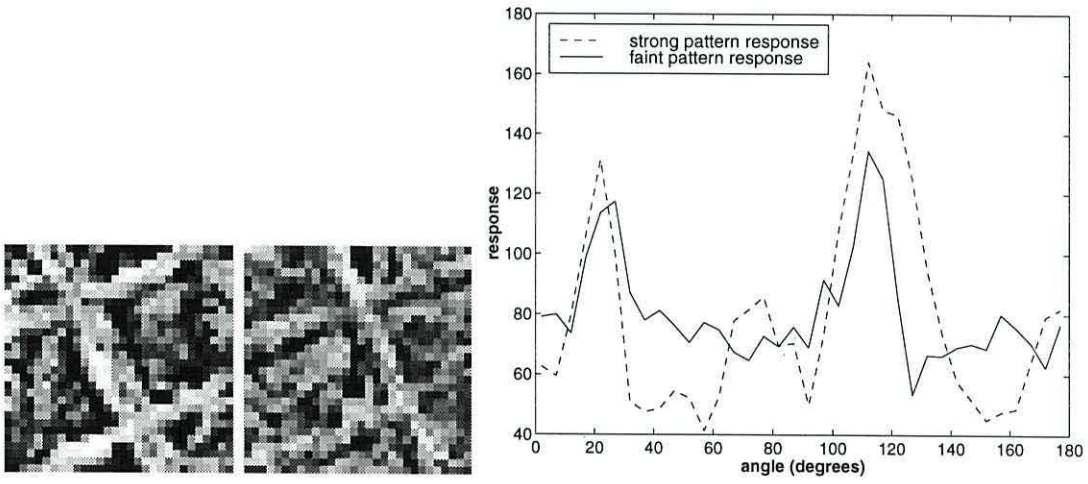


Figure 7.4: Response comparison for two skin areas of an example image showing differing absolute intensity levels of patterning in the enhanced image: The strong and faint patches analysed and the resulting profiles (CHP method).

The proper detection of similarity requires that regions showing the same patterning should result in profiles that look the same regardless of the general intensity of the region in the enhanced image. This requirement can be addressed under two differing paradigms (illustrated in figure 7.5): either all peaks in the profile deemed to be significant can be scaled so as to reach the same maximum strength or the single highest peak (and lowest trough) can be used to re-scale the profile by a constant multiplier. The disadvantages of the first (variable scaling) method are plain: firstly, the observations above concerning the multi-level nature of skin patterning and the conservation of relative proportion between skin line strength and the resulting profile mean that useful data would be lost in the reduction

of the profile to essentially "orientation only", and secondly such a process is complex compared to constant scaling and hence hard to justify given the lack of identified benefit. Using the constant scaling method the profile peak height for "primary" skin lines is kept constant over the whole image so that similar patches with similar composition of direction and relative strength result in similar profile shapes and scale.

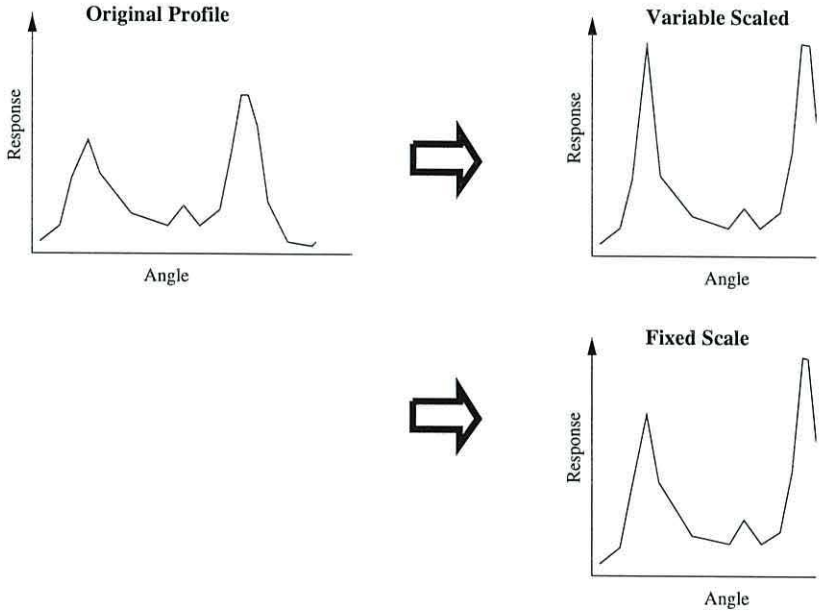


Figure 7.5: Options for the manipulation of profiles to aid in comparative measurements.

Profile fidelity to underlying skin pattern and the results of the stretching process are also affected by noise. Noise is a simple entity when described at an abstract level – it is simply the deviation of the observed from the actual. However, when examined in detail with the aims of understanding, modelling and compensation its hidden complexity becomes immediately apparent: it may be structured or random, it may be related to the data, capture device or environment or independent from them. For the purpose of this discussion the deviation in the observed profile from a representation of the "true" skin pattern character, and particularly the detection of background image variations as random weak skin lines, will be referred to as *noise*.

Although it is effective in improving the similarity of profiles for similar skin patterning, stretching does have two notable problems that need to be addressed: firstly, noise in the profile will also be magnified, and secondly (and more impor-

tantly) in cases where no significant skin lines exist (or detection fails for some other reason) the multiplier will be based on the largest peak in the background noise and as such will be unusually large resulting in the apparent detection of "random" skin lines through extreme magnification of the noise.

General noise perturbation of the profile could be reduced through the application of a smoothing filter such as a small one-dimensional gaussian, however the peaks reflecting skin lines are often quite sharp (as with the first peak in the profile graph for the strongly patterned area in figure 7.4) and hence any smoothing would cause a significant and undesirable reduction in the peak height. This adverse effect could be reduced by increasing the angular sampling frequency of the profiles, smoothing that data and then sub-sampling the result. This approach has not been followed since (as has been explained in section 6.5) the current sampling at 5° intervals is equivalent to just less than one pixel at the edge of the patch already hence no more true data is available, and in any case a higher sample rate would significantly increase the processing time especially with the addition of smoothing and sub-sampling steps.

Background noise suppression is a more pressing concern given the desire to stretch the profiles as explained above. The two methods investigated for this purpose were non-linear scaling and thresholding. The scaling approach reduces the importance of lower values and hence the low-level noise by accentuating the peaks whereas thresholding maps those values that do not rise above the profile's minimum by a large enough margin to zero. The advantage of non-linear scaling is that it does not involve a potentially sensitive parameter like the threshold, however the relative height of peaks is not maintained. The threshold–stretch combination (equation 7.1) results in profiles with good fidelity to skin pattern character whilst eliminating the problem with random orientations being presented for regions with no patterning, and is defined as,

let p be a profile,
t_n be the chosen noise threshold,
m_p, r_p be the minimum value and range of p,
and M be the required output maximum for "primary" skin lines,
then the enhancement function $e(p)$ is given by:

$$e(p) = \begin{cases} 0 & \text{if } (p(x) - m_p) < t_n \\ \frac{M}{r_p}.(p(x) - m_p) & \text{otherwise} \end{cases} \quad (7.1)$$

Figure 7.6 shows the result of applying this process with $t_n = 45$ to the profiles in figure 7.4. A threshold of 35 (in relation to a profile maximum of 255) was found to be generally suitable and is used in all the results presented later.

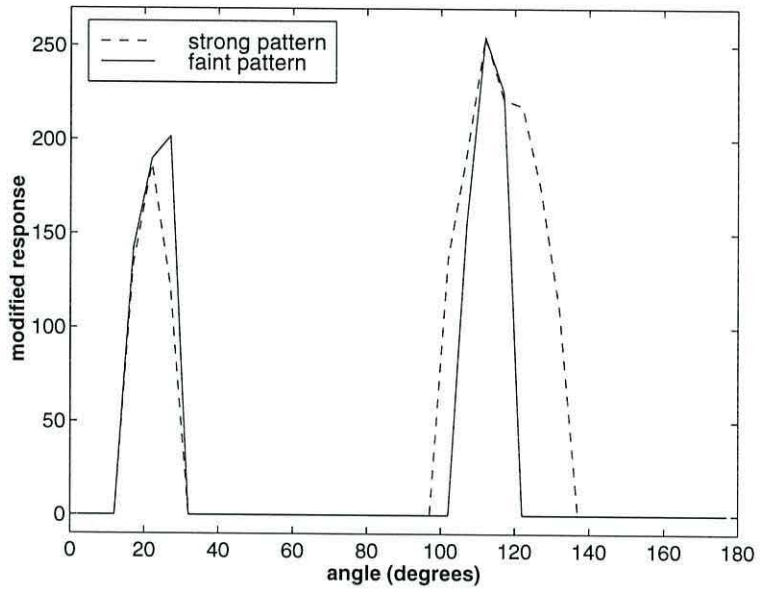


Figure 7.6: Profiles as in figure 7.4 after application of the threshold–stretch process.

7.4.2 Similarity measures

All the analysis techniques rely on comparison of profiles, either to determine the degree of consistency in a local neighbourhood or to divide the observed profiles into classes. These operations therefore require a measure which reflects the similarity of two profiles. In the initial trials standard measures were used: for the classification techniques the measure is explicit and the L2 or 'Euclidean distance' (for s element vectors p and q, $Ln(p,q) = \sqrt[n]{\sum_{i=0}^{s}(p_i - q_i)^n}$) was used, for the local variance method it is implicit in the sense that pairs of profiles are not directly compared and is related to the common statistical measure of variance.

Although the measures used were certainly effective to some degree in quantifying the similarity between profiles, consideration of the meaning of similarity when

dealing with skin patterning profiles suggests that L2 does not embody all of the salient features. Furthermore, in the case of the local variance, it is apparent that an implicit measure of similarity is unhelpful, since it is hard to explain in words exactly what a region showing a high local variance score means in terms of the local profiles. An investigation of the meaning of *similarity* when considering profiles and of possible metrics for its quantification was therefore required.

Profiles should be seen as similar when the character of the skin line patterning each represents is visually consistent. In terms of profile graphs, and with all of the observations and processes of section 7.4.1 taken into account, similarity can be expressed in terms of the following elements:

- corresponding peaks at similar angles
- corresponding peaks of similar height

The L2 measure partly satisfies both of these requirements, which perhaps explains its acceptable performance. L2 calculates the "distance" between two profiles in relation to the discrepancy between the response at each angle. Profiles with peaks of similar heights in similar locations will therefore result in a low distance.

The main failing of this measure however relates to the requirement for similarity to reflect peaks at *similar* angles. Profiles should be seen as similar even if the preferred orientations are slightly different, or in other words it is desirable that two profiles which have matching peaks except that they are shifted by $\pm$ a few degrees (and therefore by one sample in either direction) should not be seen as different. A better measure should be flexible in allowing for a small movement of this type for each peak when performing the comparison. The L2 (and Ln and $cos(\phi)$) show a rapid decrease in measured similarity where peaks do not precisely line up – the parts of both peaks not matched by the other both contributing significantly to a high distance value. With these measures, each element (sample) in the vector is independent in the sense that a global re-ordering of elements will have no effect on the calculated result. Only the natural "bleed-over" to elements which are samples of adjacent angles brings any tolerance to peak movement, and given the sharpness of the peaks, even this source of tolerance is unusually limited.

A new distance measure was therefore needed which would reflect the requirement for a flexible comparison, and therefore use the information contained in the adjacency of vector elements being responses for similar angles. The simplest way of achieving this was to view the first of the two profiles as a template and to "flex" the second to find the best comparison. Since two preferred orientations of patterned skin could alter their angle in opposite directions, the necessary flexing is more than a simple shifting of the whole second profile. Each element is compared to find the best match with one of three candidates in the first profile: either its corresponding sample, or the one either side of this. In this operation the ends of the profile are joined circularly, so that the left-hand side of the first sample is the last sample. This is sensible in terms of orientations as lines at 0° are identical to those at 180° and therefore the extreme ends of a profile are indeed adjacent in terms of orientation. The measure as described above can be expressed as in equation 7.2, however in this simple form the distance measure is flawed in two ways: Firstly, narrow peaks in the template profile can be partly "ignored" with the other profile being stretched to compare with the data either side of the peak. In the extreme case of a peak which is only one sample wide, that feature of the template profile is *completely* ignored. Secondly, the measure is not a reflexive relationship in that $d(p,q) \neq d(q,p)$ which raises the issue of which profile should be the template in each comparison and hence precludes its simple use in existing classification techniques. Both of these problems can however be resolved by the addition of the simple condition that for each sample angle, the element in the profile with the greater response is compared for best match with the three possible positions in the other, which has therefore become the template for this particular angle.

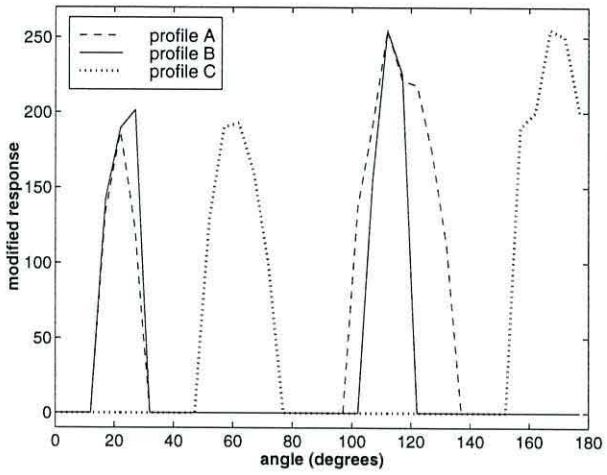
if p and q are s element vectors then,

$$d(p,q) = \sqrt{\sum_{i=0}^{s} \left[\min_{j=-1}^{1} \{(p_i - q_{i+j})^2\} \right]} \tag{7.2}$$

$$\text{CBdist}(p,q) = \sqrt{\sum_{i=0}^{s} \left[\min_{j=-1}^{1} \left\{ \begin{array}{ll} (p_i - q_{i+j})^2 & \text{if } (p(i) > q(i) \\ (q_i - p_{i+j})^2 & \text{otherwise} \end{array} \right\} \right]} \tag{7.3}$$

This measure will be called *Circularized Bungee Distance* or CB-distance and is given by equation 7.3. Figure 7.7 illustrates the effectiveness of the new CB-distance measure; profiles A and B are for patches of 'similar' skin patterning and profile C represents a different pattering, the CB-distance measure better

reflects the essential similarity of A and B (by yielding a lower distance value between them) than does L2 and yet maintains the same high distance figure as L2 when comparing the 'different' profiles A and C. Figure 7.8 shows the improved tolerance to small differences in peak centre placements for a single peak example.



Distance measures:			
L2	A-B	=	341.7
	A-C	=	834.8
CBdist	A-B	=	210.9
	A-C	=	834.8

Figure 7.7: Improvement in quantifying profile similarity: Three profiles with distance calculations using CBdist and L2 – Profiles A and B form a "similar" pair and profiles A and C form a "non-similar" pair.

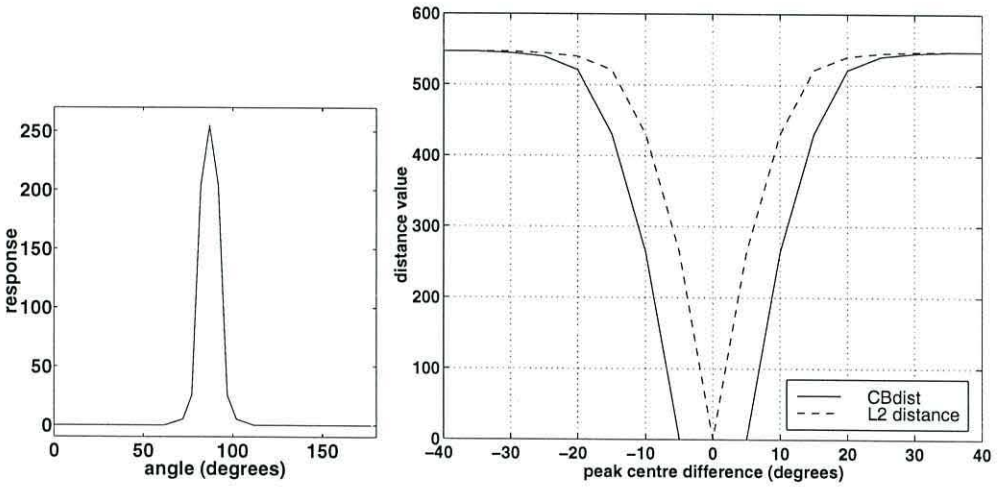


Figure 7.8: Variation in distance measure values with displacement of a single peak in a profile: CBdist showing improved tolerance to small shifts over L2.

7.4.3 Local Variability Analysis and Evaluation

As remarked in section 7.4.2, the use of an implicit measure of similarity in the local variance method is unhelpful when it comes to explaining in words

exactly what the local variance figure means. This consideration, together with the availability of the new CB-distance measure, prompted the adoption of a new method of evaluation to replace 'local variance'.

The new metric still quantifies the local variability in a 3x3 window centred on the current patch, but the value is obtained by taking the median of the CB-distances from the profile of the central patch to those of each of the eight surrounding patches. The median (rather than mean) value is taken to avoid undue effects of isolated unusual patches in the window and to improve the performance (reduce blurring) at the edge of a consistent area. Where all eight neighbours are not available (at the edges of the profile image) just the data that is available is used so that unlike for the local variance implementation the values at the edge of the local variability image can be used. The information given by these edge results is naturally less reliable, being based on less information (although, except in the corners, more than half of the neighbours are always available). This edge data is used in spite of these problems as it still provides valuable information on local similarity in the skin area, especially in the sample image set used where there is not always a great deal of skin area surrounding the lesion. Figure 7.9 compares this new local variation result to that of the original local variance method on the artificial lines image.

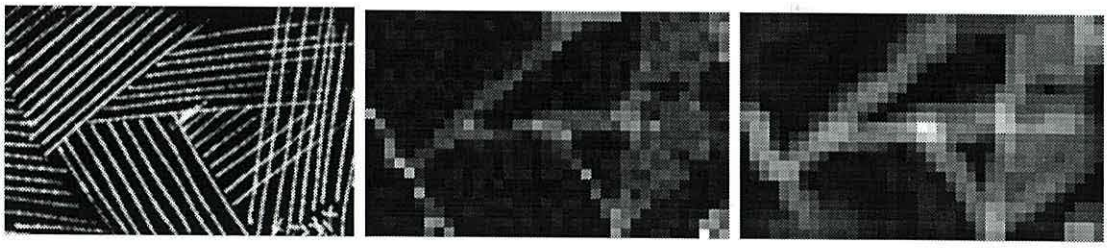
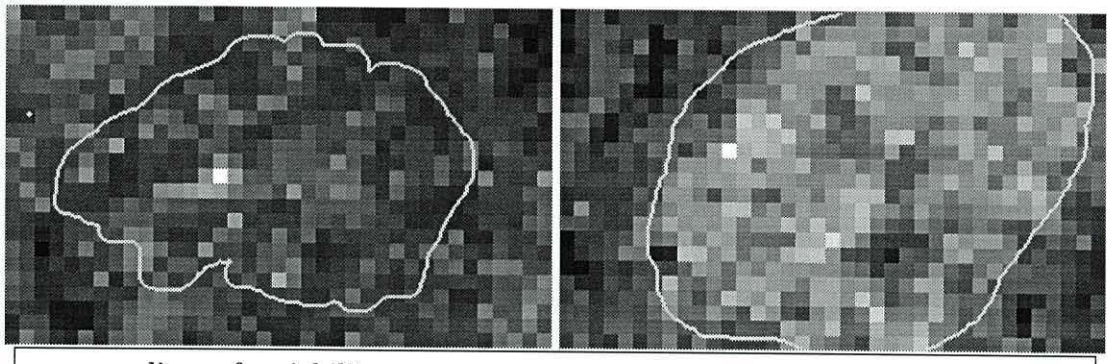


Figure 7.9: Local variation measures on test data: Test image (left) with new method (centre) showing reduced blurring at the boundaries and increased stability within consistent regions over the local variance method (right).

Evaluation of disruption is performed in a similar fashion to that used for the 'local variance' except that the median and difference are used rather than mean and ratio. The median value for the disruption (local CB-distance median) is calculated for both the skin (m_S) and the lesion (m_L) areas and the difference $m_L - m_S$ then represents the contrast between the areas. A high value indicates that the patterning is significantly less consistent over the lesion than on the

surrounding skin, and hence that a high level of disruption has been caused by the lesion. Figure 7.10 shows the new local variability data overlayed with the lesion extent together with the results for the area medians and the final difference result.



area medians of variability:

naevus (left)	lesion	=	51.8	melanoma (right)	lesion	=	88.2
	skin	=	49.3		skin	=	50.3
	difference	=	2.5		difference	=	37.9

Figure 7.10: New local variability measure on real data with overlay of lesion extent: A benign lesion (left) showing consistency and a malignant lesion (right) showing marked increase in variability in the lesion area.

The median value for the skin and lesion regions is used in preference to the mean in order to remove the effect of anomalous profiles in regions of otherwise consistent patterning. In practice little difference between the mean and median values is observed, although the median value is generally a little lower and where profiles are extremely consistent the median is indeed seen to be more faithful. The final disruption figure therefore shows very little change through the adoption of the median except in cases where the profiles are generally highly similar except for a few anomalies.

The difference is used rather than the ratio for two main reasons: Firstly, the CB-distance values reflect a constant scale of similarity so that two profiles p_1 and p_2 with CB-distance d are as 'similar' as q_1 and q_2 with CB-distance d, hence where the lesion and skin regions have similar medians for their local similarity, the lesion should not be seen as having caused significant disruption. The problem with using the ratio measure is that as the median for the skin region rises, the same disruption measure results for lesion areas reflecting increasingly less 'similarity' to the skin. This effect is highly significant where the skin area shows

an unusually low variability, in which case, a small increase over the lesion area is reflected in an exceptionally large disruption figure. Secondly, use of difference has the added advantage that the disruption figure retains a clear relationship to the values obtained from the distance measure.

7.4.4 Using CB-distance in Classifiers

The increased effectiveness of the CB-distance measure in relation to quantifying the important elements of similarity in profiles can also be employed in the classifiers, mRAC and SOFM.

mRAC The adoption of the CHP profiling method together with the noise reduction and profile enhancement measures as discussed, and the new CB-distance measure requires a re-consideration of the parameter set for the mRAC segmentation.

All parameters were kept constant for all the image and disruption evaluation results given below. Trial values for the parameters were set through consideration of CB-distance values between 'similar' and 'different' profiles (such as those displayed in figure 7.7). The splitting t_s and conservative merge t_{cm} parameters were again found to be non-critical where they are set with reference to the goals of the mRAC stages they control: t_s should cause the split of all but the most uniform regions and then t_{cm} should merge only those regions that are highly uniform. The setting of t_{optm} obviously has the greatest effect on the final segmentation. The thresholds used for all the following results are; splitting $t_s \approx 30$, conservative merge $t_{cm} \approx 60$ and optimal merge $t_{optm} \approx 306$.

Preliminary trials using the new data indicated problems with the validity of the cleaning stage of the process (which force-merges small regions to facilitate interpretation by simplifying the segmentation) and ultimately resulted in it being removed. When using the CHP method, many profiles in highly disrupted regions have no neighbour with a similar profile (a result of the increased sensitivity and fidelity of the method over autocorrelation). Since in such cases a profile may *legitimately* have no neighbouring partner profile (a singleton class), force-merging

would create false regions rather than clean the segmentation by merging isolated anomalies into large neighbouring regions as intended.

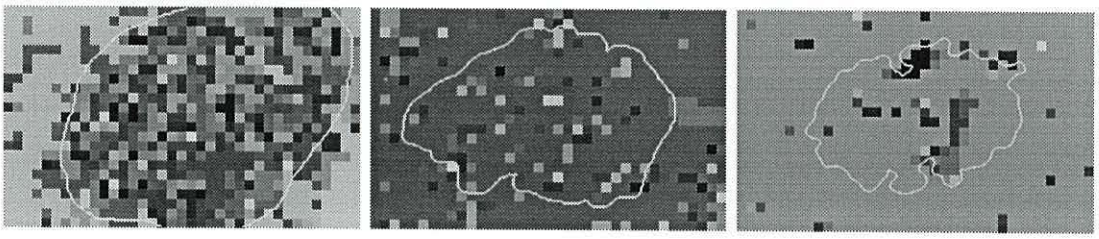


Figure 7.11: mRAC classification results (no cleaning) using the CHP data and the CB-distance measure for similarity with overlay of lesion extent: The classification for the melanoma (left) reveals the lesion whereas that for the two naevi (centre and right) does not.

Figure 7.11 shows the mRAC classification results using the CHP data with noise reduction and normalization and the CB-distance measure for similarity. The melanoma (left) shows a classification which reveals the lesion as a multitude of separate classes within surrounding skin which requires much fewer classes. This situation is in contrast to the two examples of naevi which have no obvious correlation between the classification and the overlayed lesion extent and a similar fragmentation of the lesion and skin areas.

SOFM This classifier partitions the data into a fixed number of classes and hence the analysis of the processed CHP data data does not require a re-evaluation of a parameter set. The CB-distance measure was used in place of L2 in all the SOFM results below.

7.4.5 Unifying Classifier Evaluations

In the initial trials a separate evaluation technique was devised for each of the methods of analysing the profile data. This allowed the particular qualities of each to be used in the calculation of the final disruption metric. With the exception of the local variation methods the analysis of the profile data begins with a classification stage where the profiles are grouped based on their similarity. The ability to compare the effectiveness of these classifiers on the profile data would obviously be of value, however since there is no obvious 'correct classifi-

cation' a quantitative comparison at this level is not available. Since the aim of the classification is to allow the quantification of skin patterning disruption, the performance of the various classifiers is then best evaluated in terms of the suitability of the disruption results it yields. Unfortunately, the use of individual and *ad-hoc* evaluations for each effectively introduces another unknown parameter preventing the results from being a valid comparison of the classifier. In addition, the introduction of new classifiers is not only hindered by this difficulty in comparative assessment but also complicated by the need for the development of an associated evaluation method. This situation of pairings of classifier and dedicated evaluation is undesirable then in terms of both testing and future development, a single method for evaluating any classification of profiles in relation to lesion/non-lesion regions was therefore investigated.

The main premise for the skin patterning work is the existence of consistency in preferred directions in skin patterning, consequently the skin area of a lesion image should be reflected in profiles that show similarity over extended regions. Effective classifiers should then divide the majority of the skin area into a relatively small number of classes (a few for legitimate skin pattern and some anomaly classes). If a lesion is not disrupting the skin pattern then the same situation will continue there with the legitimate skin pattern classes also appearing over the majority of the lesion. The existing evaluation methods for mRAC and SOFM classifications both use the concept of a *true-skin* class where a class is designated true-skin if it accounts for more than than $t_{ts}\%$ of the skin area. All the true-skin classes taken together then form a 'model' for the local skin against which the lesion area can be compared.

However, the true-skin classes cannot be effectively identified using the same value for t_{ts} due to the differences in the mechanics of the classifiers. mRAC creates classes by merging regions in order of decreasing similarity until a threshold is reached, thus the number of classes and their sizes are unconstrained whereas in SOFM the number of classes is fixed and each must have a similar size. In practice this means that a relatively large number of smaller classes must be designated as true-skin for SOFM and hence a much lower value of t_{ts} is needed.

Since a single setting for t_{ts} will be ineffective for all classifiers a common eval-

uation method needs to encapsulate the essence of true-skin without recourse to such a threshold parameter. Sorting the classes by their % contribution to the skin area and then plotting the cumulative % coverage of the skin and lesion areas allows a more detailed investigation of the character of the classification with respect to the skin and lesion regions (% coverage is used as it accounts for differences in the relative size of the lesion in the image). Where the classes have comparable significance in the skin and lesion areas the lesion coverage will rise at a comparable rate to that of the skin, whereas when separate classes are being used for the lesion, area coverage will lag behind that of the skin until later. Figure 7.12 shows model and real examples for such contribution graphs for malignant and benign lesions and the 'area' which represents the degree of distinctness of classes in the lesion area.

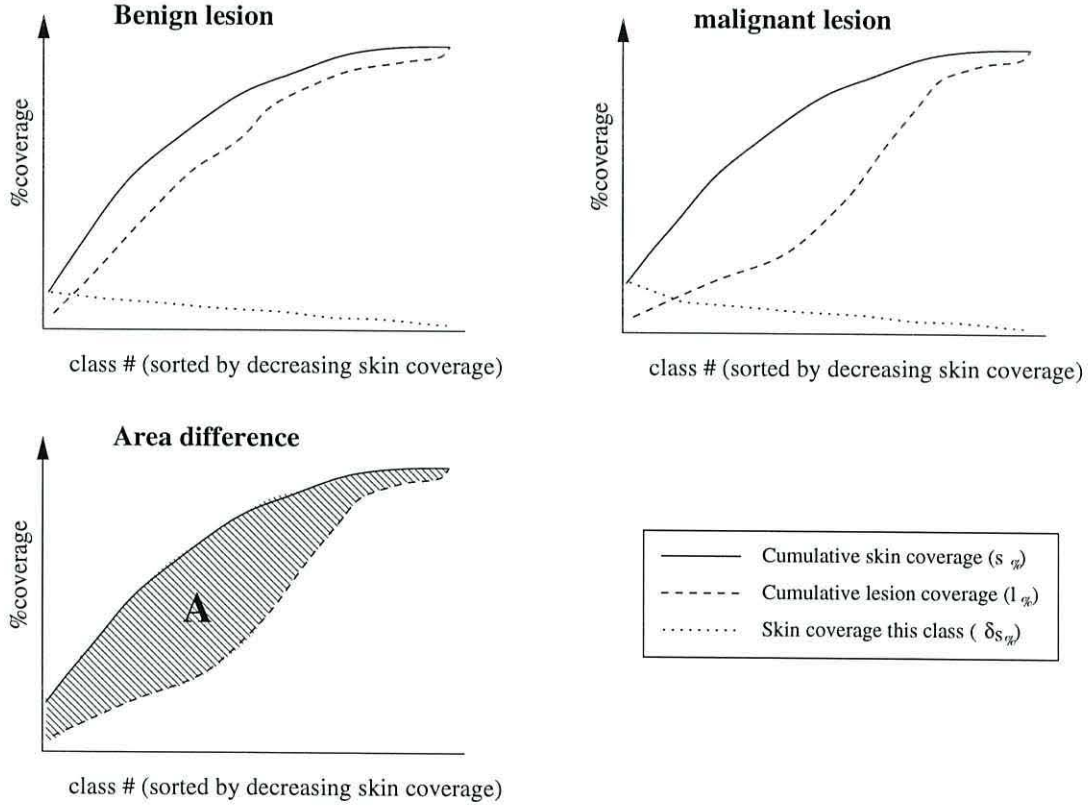


Figure 7.12: Area difference concept based on true-skin classes: top: conceptual graphs for a benign lesion (left) and malignant lesion (right) and bottom: illustration of 'area' A and legend.

In terms of such graphs the original mRAC evaluation is simply 100-$l_{\%}$ at the point where $\delta_{s\%}$ falls below t_{ts} and the original SOFM evaluation is the difference between the class number at that point (for a different t_{ts}) and the total number

of classes which is a constant. The use of area A as the evaluation measure has the desired effect of removing the dependency on t_{ts} and unifying the evaluations, however there are two major factors affecting this area A evaluation: Firstly, it is unduly affected by differing numbers of classes as it takes no account of the importance (overall size) of each class. Secondly, the difference resulting from an early class in the sorted list in effect persists from that point on and is 'recounted' due to the cumulative nature of $l_{\%}$ and $s_{\%}$. Simple area evaluations such as $A = \sum_{\forall \text{ classes in order}} |s_{\%} - l_{\%}|$ performed on a class-by-class basis will suffer from both these factors. The effect of the first can be removed through the addition of class size weighting (in terms of the graphical representation this means replacing class number on the x-axis with cumulative coverage of the whole image area, $w_{\%}$). Both factors can be dealt with though if the individual class contributions are considered rather than the cumulative figure, the calculation then takes the form $B = \sum_{\forall \text{ classes}} |\delta_{s\%} - \delta_{l\%}|$, the sum of the absolute differences between the contributions of each individual class to the skin and lesion areas. Since this calculation is no longer based on cumulative figures the ordering of the classes is not important, however if the classes are still ordered by skin coverage, the information can be plotted meaningfully as in figure 7.13 which shows characteristic graphs for malignant and benign cases and the new area B. In practice the contributions of each class to the lesion area can vary quite widely, however for the purpose of illustration a smoothed mean line is shown.

Viewed from a different perspective the evaluation of the classification can be approached using the statistical technique of χ^2 likelihood. The χ^2 statistic measures the probability of there being a relationship between two different partitions of a set. The individual profiles form the set of observations, the lesion and skin areas form a primary classification for this set and the classifier result becomes a secondary classification. There are two main formulations for the statistic as given in equations 7.4 and 7.5 although both are essentially similar. Pearson's χ^2_p (eqn. 7.4) will be used in this discussion.

$$\chi^2_p = \sum_{i,j=0}^{i=n_1,j=n_2} \frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}} \qquad \text{(Pearson's)} \quad (7.4)$$

$$\chi^2_l = -2 \sum_{i,j=0}^{i=n_1,j=n_2} O_{i,j} \ln \frac{O_{i,j}}{E_{i,j}} \qquad \text{(Likelihood)} \quad (7.5)$$

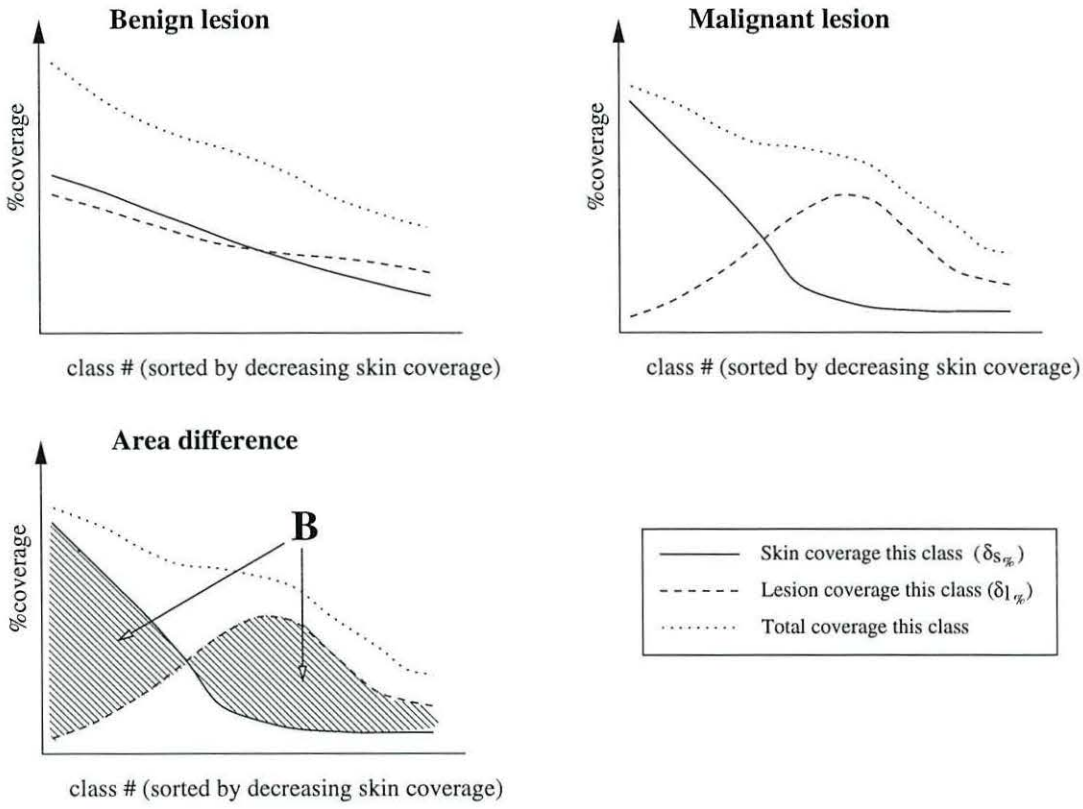


Figure 7.13: Individual class contribution area difference: top: conceptual graphs for a benign lesion (left) and malignant lesion (right) and bottom: illustration of 'area' B and legend.

Where: The primary and secondary partitions are of n_1 and n_2 classes respectively

and $O_{i,j}$ is the number of actual observations in primary partition i and secondary j

and $E_{i,j}$ is the expected number of observations in primary partition i and secondary j given the total number of observations primary i and the total in secondary j

In essence the χ^2 statistic measures the strength of evidence against a hypothesis that the primary classification gives no information on the secondary. Under this hypothesis, the division of the observations in class j of the secondary classification between each class in the primary will be in exact proportion to the size of each of the primary classes. Any deviation from this expected pattern is evidence against the hypothesis and therefore shows a connection between the classifications. The

strength of evidence for such a connection can be evaluated as: $C = \sqrt{\frac{\chi_p^2}{\chi_p^2+T}}$ where T is the total number of observations. It is recognized that there are problems with this measure of strength, but it is widely used and provides useful information.

The χ^2 statistic is in some ways similar to the evaluation of area B as described above where the deviation from an expected value for each (secondary) class is evidence for the skin and lesion areas not following the same pattern of classification. The area B calculation assumes that if there is no relationship between the classification and the skin/lesion partition then the portion of these areas covered by each class will be the same, whereas χ^2 assumes that the skin and lesion area coverage for a given class would reflect the size of the skin and lesion regions respectively.

The assumptions involved in both the area B and χ^2 calculations have two significant problems in terms of their use as evaluations: Firstly, regardless of the classification technique used, singleton classes cannot satisfy the assumption for either method. A singleton class in the skin area for example will obviously not represent a similar portion of the lesion area and the skin and lesion area coverage will be greater and less than the expected values respectively; even if there was no real relationship, singleton classes will always indicate that there is. Taken to the extreme situation where all the classes are singletons, both the area B and χ^2 evaluations will report a maximum evidence for skin and lesion area difference, whereas in truth the segmentation is providing no information at all. All small classes in fact have a similar problem similar to that of singletons, in that any small sample from a random process can be misleading, related to this situation, a class containing only three observations could easily fall only within the skin region simply by chance. Ignoring all very small classes would of course offer a solution to these problems providing that only a small number exist so that the impact of the 'holes' in the classification data remains small, however the small class problem only becomes important where there are many such classes, and in that situation they could not be ignored.

The second problem with the assumptions is particular to the mRAC classification. The region based property (regions forming by local merging) of the mRAC

process means that the final classification is inherently regional. This is a valuable property given the aim of identifying local similarity between profiles and that adjacency does entail an increased likelihood of two regions having similar skin patterning. It is however necessarily in conflict with any assumption that a single class would ever have a similar representation in different areas of the image. Given a random matrix of profiles most regions are likely to be small (since no chain of similar profiles is likely to extend very far) and any class (which is in fact a small region) will therefore be most likely confined to one of the skin or lesion areas. In practice, the skin pattern is often both sufficiently consistent and well detected that large regions will form to represent the skin patterned area (connecting chains of profiles are indeed available) and these regions either extend over the lesion area if the pattern is not disrupted, or they stop close to the boundary between the lesion and the skin. However the underlying conflict for the mRAC classification remains and does indeed appear to affect the consistency of the evaluation.

These observations indicate that a different evaluation suited to the properties of the mRAC classification is required. The desire to remove the dependence on a threshold still remains, consequently a return to the original true-skin based evaluation is not desirable. With reference to the properties of effective classifications described at the beginning of this section an alternative comparison for the skin and lesion areas can be formulated based on the difference in *fragmentation* of the two areas. Skin patterned areas will require a small number of classes when compared to areas where this pattern has been disrupted. The fragmentation of an area can be defined as the number of classes used for that area as a fraction of the number of profiles in the area (equation 7.6). Disrupted lesion areas are then expected to have a higher fragmentation than the surrounding skin and a comparison δ_{frag} can be obtained by simple subtraction of the fragmentation results (equation 7.7).

For an area A of a matrix of profiles and a classification C the fragmentation is defined as:

$$f_C(A) = \frac{\#\text{classes appearing in A}}{\#\text{profiles in A}} \tag{7.6}$$

and, for lesion area L and skin area S:

$$\delta_{frag} = f_C(L) - f_C(S) \tag{7.7}$$

Such an evaluation is obviously most suitable where the number of classes is variable as a fixed number of classes will cause a certain level of fragmentation even if all profiles were identical. The fragmentation measure is particularly suited to mRAC as it is a regional measure which makes no assumptions based on the distribution of class membership. In addition the measure works well even where the number of classes is large and the advent of many singleton (or small) classes has no particular adverse effects on the result. These are both important since the regional nature of the mRAC process and the sensitivity of the new CHP profiling often results in relatively large numbers of small classes, particularly in disrupted lesion areas (see figure 7.11).

In conclusion then, both the area B evaluation and χ^2 C strength are useful measures for comparing classifications given knowledge of the extent of the skin and lesion areas. They would allow for comparison of the effectiveness of different classification techniques on the profile data given that these techniques are equally consistent with the required assumption that classes in a 'random' case will have representation in skin and lesion regions in proportion to the size of these and that the problems of singleton/small classes are monitored or addressed. Neither of these techniques are however suitable for mRAC classifications, and the inherent nature of mRAC suggests instead the use of a measure based on difference in fragmentation.

7.4.6 Re-considering ART

The introduction of the CHP profiling method when taken together with the noise reduction and profile enhancement steps and the new CB-distance measure

prompts a re-consideration of the ART classifier. One of the biggest problems with ART had been identified as its built-in normalization of profiles prior to classification. This feature when used in conjunction with less well detected skin patterning resulting from the autocorrelation method produced unusual classifications. Profiles that were essentially flat, indicating no clear detection of any skin patterning, were being stretched so that random noise was being identified as real patterning. Not only would the increased sensitivity and fidelity of the CHP profiling method reduce such problems, but the noise reduction and profile enhancement steps themselves perform normalization so that the effects of the internal normalization in ART are be significantly reduced.

However it is important to recognise that the two methods of normalization are fundamentally different and as such the effects of the internal manipulation in ART cannot be ignored. The linear stretch normalization employed for profile enhancement scales each profile so that the largest peak is always of the same strength, consistent with the assumption that the strongest skin line in each patch represents the primary lines in the skin pattern. The ART classifier however, normalizes each input to a unit vector. Such processing introduces dependency on the overall level of each profile, such dependency is undesirable as illustrated by the following two examples. Firstly, given two profiles that are essentially identical in terms of the preferred orientations which they represent except that one shows wider peaks for each orientation, the latter will have all the peaks lowered, thus falsely increasing the calculated difference between them. Secondly, given two profiles representing the same preferred orientation, if the second also shows another orientation, this profile will be lowered with respect to the first again falsely increasing the calculated distance between them.

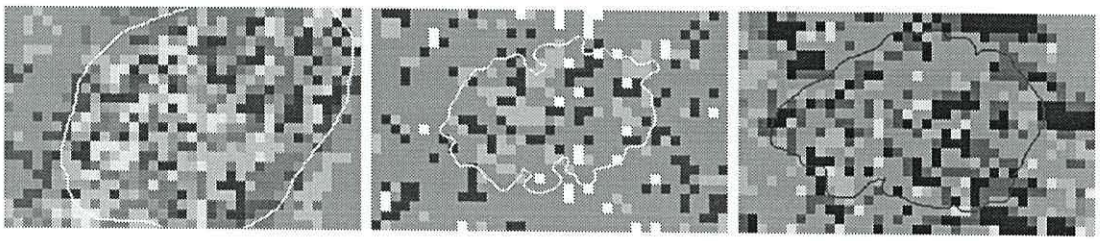


Figure 7.14: ART classification results using the CHP data and the CB-distance measure for similarity with overlay of lesion extent: The classification for the melanoma (left) reveals the lesion whereas that for the two naevi (centre and right) does not.

In spite of these problems ART produces meaningful classifications of the profile data as illustrated in figure 7.14 which shows the melanoma (left) revealed by the classification whereas the two benign naevi show similar structure and classes throughout the image.

Each of the profiles extracted from the image are presented to the ART classifier in random order. For each profile, p, the most similar of the exemplar profiles for all classes created so far is found. If this class sufficiently well represents p then the exemplar is modified so as to be slightly closer to p, if not then a new class is created with p as exemplar. The complete set of profiles are presented a number of times to allow the classes to stabilize. The two main parameters which control this process are the threshold deciding sufficient similarity (vigilance) and the modification (learning) rate. The number of presentations can also significantly alter the result.

The vigilance parameter cannot be set directly in relation to empirical evidence of similarities as in mRAC due to the effect of the unit vector normalization, however a value of 0.8 relates to a CB-distance of around 300 (as used for mRAC) if a simple scaling to range 0-1 is assumed. The examples in figure 7.14 use a vigilance of 0.8 with a learning rate of 0.01 and 20 presentations. A slightly improved balance between division of disrupted areas and unification of consistent regions was achieved using a vigilance of 0.795 and this value is used for the results presented later.

Although the ART classifications are generally effective, the results for another run of the same data are not always consistent even with all parameters unchanged. The random presentation of the data means that the original exemplar for each class is likely to be different for each run. This in turn means that given the early presentation of all the extreme examples of profiles which would normally fall into a single class (as in an undisrupted skin area), the result is the development of several classes essentially representing the same profile.

7.5 Final Results and Discussion

This section shows the results obtained using the developments on the analysis–evaluation process described on an expanded set of 22 trial images from the same source as for the initial trials. This is a superset of that used before and contains all but one of the melanoma examples available in the 54 image set (one had to be excluded since there was no significant skin area surrounding the lesion). The set contains 8 melanomas and 14 naevi (either compound or junctional). The melanomas are of various types, sizes and stages of development and the naevi include dysplastic and atypical examples.

The final results are presented separately for each of the evaluation techniques using two main forms. The first shows the results separately for the individual histological diagnoses, and the second compares the results against a visual estimate for the disruption in the skin patterning. This second comparison is vital since the aim is to demonstrate that the techniques are able to quantify the patterning disruption so that the performance should not just be measured in terms of discriminant ability on lesion histology. A high correlation between the result and the visual disruption estimate indicates an effective acquisition, analysis and evaluation sequence.

Finally a Receiver Operating Characteristic (ROC) curve is also shown for each evaluation. An ROC curve [91, 92] gives an indication of the effectiveness of a continuous measure in discriminating between two classes of example data: The greater the area under the curve, the more successful the measure is. The theory underlying ROC curves is as follows. Given a set of examples (e.g. images) divided into two classes, 'positive' and 'negative', the accuracy of a binary indicator can be illustrated in the form of a table (figure 7.15). The sensitivity of the indicator

		Truth	
		positive	negative
Indicator	positive	a	b
	negative	c	d

Figure 7.15: Final results on a larger data set.

is then defined as $\frac{a}{a+c}$ and the specificity as $\frac{d}{b+d}$; a useful indicator should have high sensitivity and specificity. Many indicators (such as those described here) can take on a continuous range of values indicating a degree of 'positiveness'. For such indicators each choice of threshold value produces new table, and a value for the sensitivity and specificity. A plot of sensitivity (true positive fraction) against 1-specificity (false positive fraction) is a ROC curve. The diagonal line represents chance whereas a curve that is well above the diagonal shows an accurate indicator and will have an area greater than 0.5.

7.5.1 Local Variability Measure Results

The first set of results are for the new local variability measure and are shown in figure 7.16 (ROC curve and data in figure 7.19). A good correspondence with the visual estimate is achieved with a correlation coefficient of ≈0.79, and performance with respect to the histological diagnosis is also generally good.

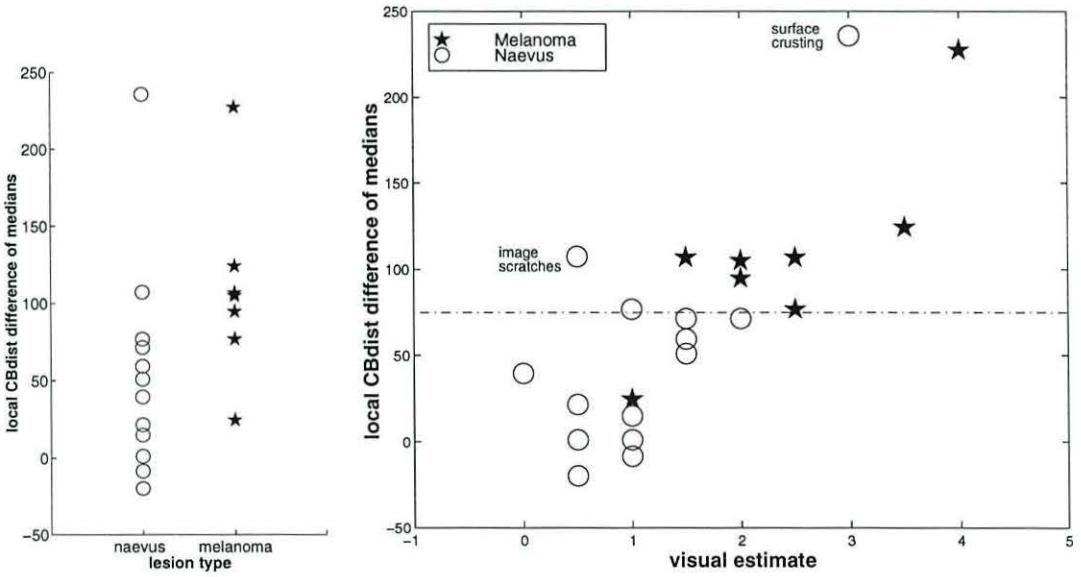


Figure 7.16: Final local variability metric results on a larger data set.

An interesting example of the problems in evaluation of feature detection in relation to lesion diagnosis is provided by the one melanoma which apparently produces an exceptionally low result (this example produces a low result under

all the methods). Although the lesion is cancerous, its low disruption result is *not* anomalous; the visual estimate of patterning disruption is also very low indicating that the automatic estimate is accurately reflecting the *feature* if not the *diagnosis*. The melanoma is small and thin (0.4mm) and relatively early in its development (Clark level II) which might explain its minimal interference with the skin pattern. Two of the example naevi however produce truly unusual results. The first of these (labelled 'surface crusting' on the scatter graph in figure 7.16) appears as the highest disruption value for all the examples and is an apparent failure in terms of diagnostic performance, however the estimated skin patterning disruption figure is also unusually high (for a naevus) so that the result does in fact show a valid detection of disruption. The image itself shows distinct crusting over the entire lesion area and a consequent large scale loss of patterning consistency (figure 7.17). Such an example would be expected to result in a high disruption figure.

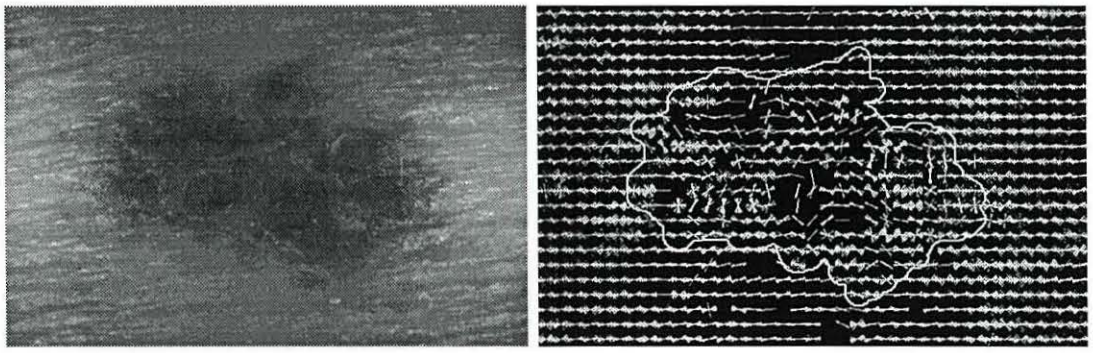


Figure 7.17: Unusual example of a benign naevi: Original image (left) showing distinct crusting over the lesion area and illustration of the detected patterning overlayed with the lesion extent (right).

The second (labelled 'image scratches') is most unusual in its deviation from a normal correlation with the estimated disruption. The source of this deviation can be traced to the presence of fine 'scratches' in the original image mostly over the lesion area (see figure 7.18 top left). The scratches are seen as a bright central line with a dark line to either side (the latter being especially apparent in the darker lesion area), presumably the result of the pigment on the slide being removed from the scratch line and pushed to either side. These features have a double impact on the profile detection with the pattern lines of the preferred direction being completely broken by the scratches and the concentration of pigment either

side forming continuous lines and therefore false skin pattern lines at a completely different orientation. This results in a high local variability result in the region of the scratches, and since they are concentrated in the lesion area (and besides have more impact there) the lesion area is found to have relatively low consistency of pattern compared to the surrounding skin and consequently a high pattern disruption result is reported. Figure 7.18 shows the original image and below it a second, manually altered ('touched up') version. The local variation results (to the right of the images) show bright (high local disruption) areas corresponding to the scratches indicated in the original and a distinct reduction in these for the altered image. The final disruption result for the altered image is 75.0 compared to 107.4 for the original. The new result better reflects the low visual disruption estimate although it is still a little high as a result of the remaining scratches.

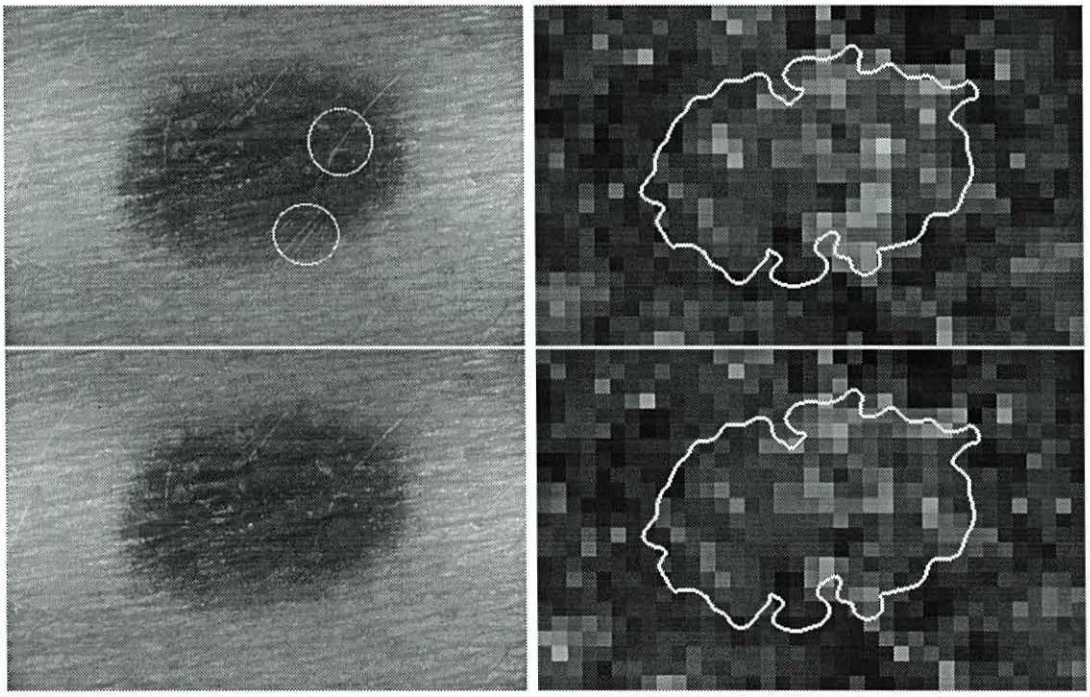


Figure 7.18: Image 'scratches' and their effect on the local variability measure: Top: Original image (left) with prominent scratches circled and local variability (right) showing corresponding high values. Bottom: Manually 'touched-up' version (left) showing much reduced local variability results in the two edited regions (right).

Is is important to note that the presence of a few hairs in the original image does not affect the local variation results in the same way or to the same degree as

the scratches described above. The problem with hairs in the image is again in their effect at the skin patterning enhancement stage, the hair is highlighted as a strong skin line and consequently is detected by the profiling stage, adding a false preferred direction to the profile at that point (unless the hair falls in line with the true direction). The key difference is that the hair *adds* to the underlying pattern rather than replacing it as with the scratches. The effect on the disruption result is minimized by two elements of the local variability measure: Firstly, since the true pattern profile is still detected despite the presence of the hair and the hair therefore only adds an extra component to the profiles it influences, the profiles still retain significant similarity. Secondly, a hair will commonly only affect a small number of profiles in the measure's local window. The use of the median of the eight local distances then means that affected profiles will often have no influence on the final result.

In spite of both of the unusual naevus results the ROC analysis (figure 7.19) still indicates that the measure is effective as a discriminator between the melanoma and naevi groups. With these two unusual examples removed the performance is extremely good with a ROC area and correlation coefficient both around 0.9.

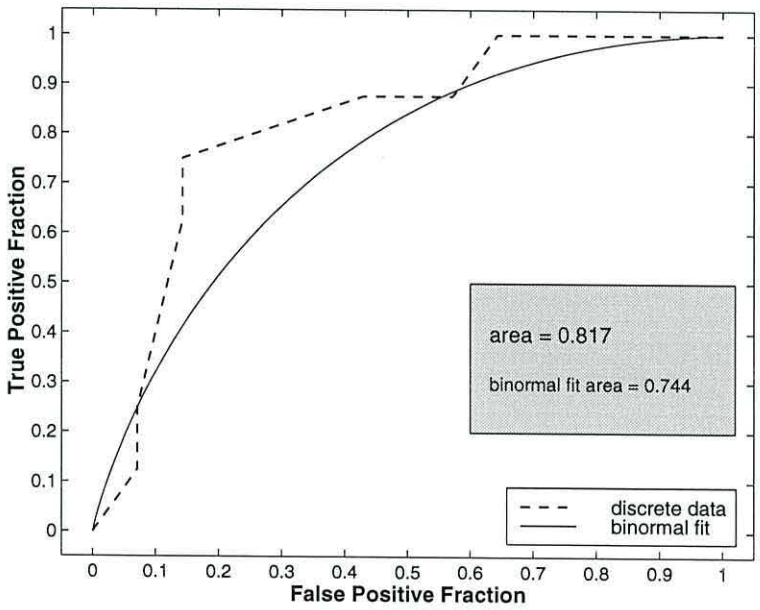


Figure 7.19: ROC curve and analysis for the final local variability metric results on the larger data set.

7.5.2 mRAC Fragmentation Results

The second set of results are for the fragmentation difference measure based on the mRAC classification using the CB-distance technique for the similarity criterion and are shown in figure 7.20 (ROC curve and data in figure 7.22). The correspondence with the visual estimate is slightly improved over the local variation measure and has a correlation coefficient of ≈0.86. Performance with respect to the histological diagnosis is also better with a clearer separation between the two categories.

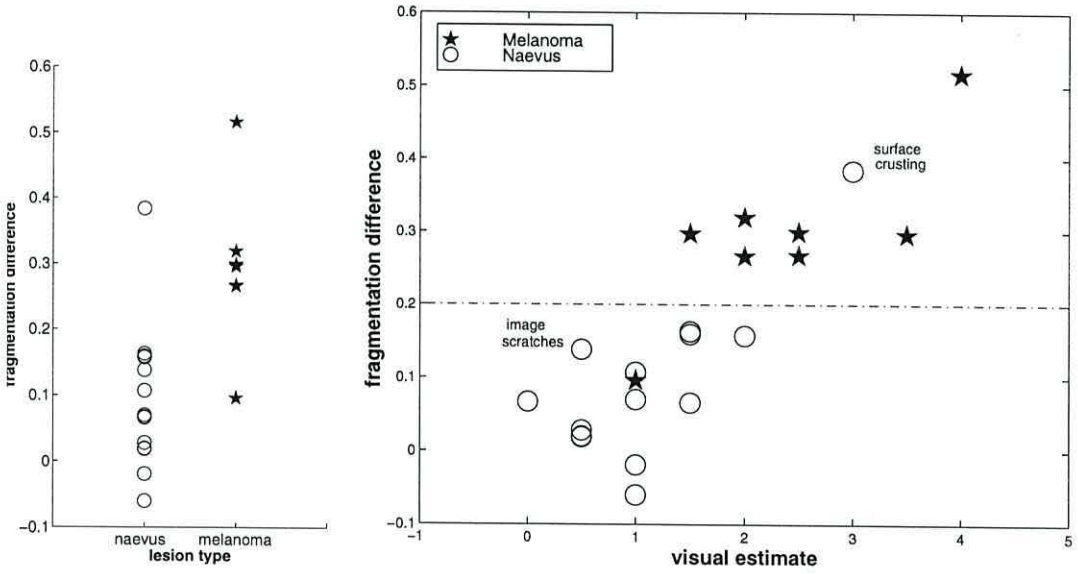


Figure 7.20: Final mRAC fragmentation evaluation results on a larger data set.

The two naevi examples which caused the unusual results when using the local variability measure are identified in the new scatter graph of figure 7.20. The example showing surface crusting again has a high figure for disruption although it is no longer the highest result and corresponds better with the visual estimate. The example with image scratches produces a less abnormal result under this measure although it is still arguably higher than expected. The locality of the scratches is classified separately from the normal skin pattern as would be expected, however the majority of the skin and lesion area are represented by a single class in spite of the scratches (figure 7.21 left). The presence of such large classes in both skin and lesion areas ensures that the fragmentation results are relatively low for both areas and hence that the difference is small. In addition,

the scratches themselves provide some degree of similarity so that scratch influenced profiles are not always forced into singleton classes (figure 7.21 left) and hence the effect on the fragmentation result is reduced when compared to the presence of random pattern disruption. The scratches do still have a degree of influence though with the manually touched-up version of the image resulting in a cleaner classification in the lesion area (figure 7.21 right) and a fragmentation difference reduced from ≈0.14 to ≈0.09.

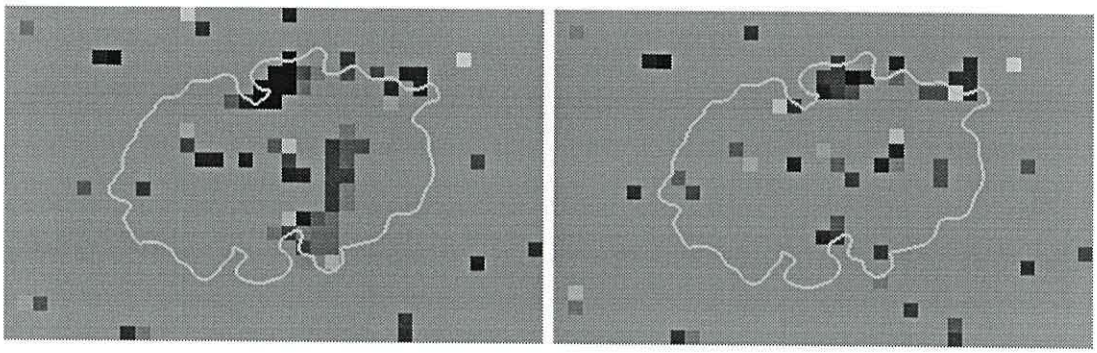


Figure 7.21: Image 'scratches' and their effect on the mRAC classification: Classification for the original image (left) showing greater fragmentation in the lesion area when compared that for the manually 'touched-up' version (right).

The ROC analysis for the mRAC fragmentation evaluation (figure 7.22) indicates that the measure is again effective as a discriminator between the melanoma and naevi groups and that its performance is slightly better than the local variation measure on the full set of data. For the unusual naevus example with the surface crusting removed or re-classified as a melanoma, the ROC area becomes greater than 0.95.

7.5.3 SOFM Results

The third set of results are for the SOFM classification again using the CB-distance technique for the similarity criterion. The χ^2 results are shown in figure 7.23 (ROC curve and data in figure 7.24). The correspondence with the visual estimate is reasonable although the spread is greater than with the local variation and mRAC fragmentation measures and has a correlation coefficient of only ≈0.55. Performance with respect to the histological diagnosis is unimpressive with no clear separation between the two lesion categories.

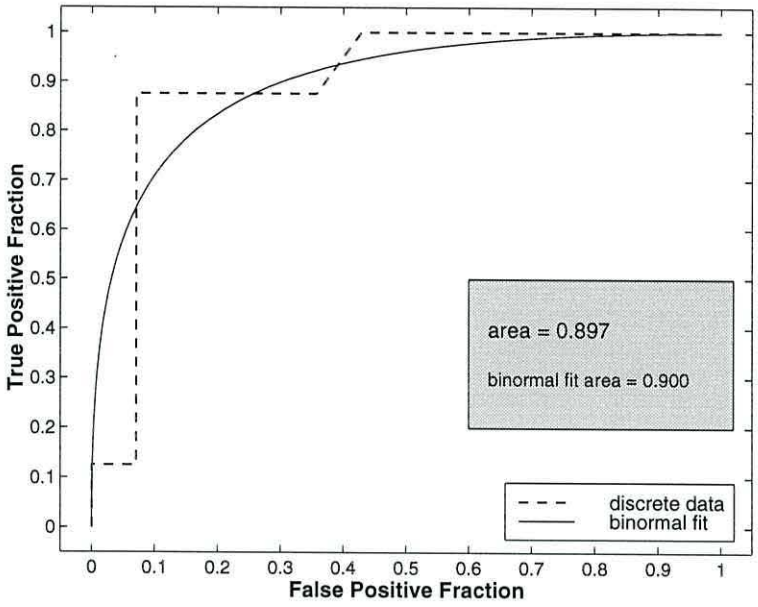


Figure 7.22: ROC curve and analysis for the final mRAC fragmentation metric results on the larger data set.

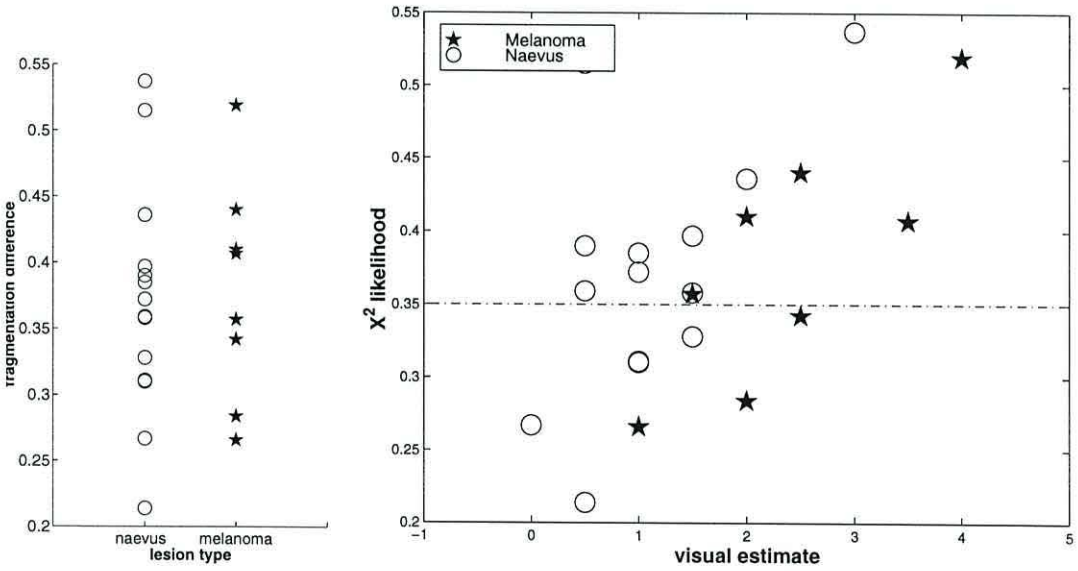


Figure 7.23: Final SOFM χ^2 evaluation results on the larger data set.

The ROC curve for this data shows clearly the poor performance with respect to discrimination between the lesion categories with an area of $\approx$0.53, only slightly better than chance. Although the performance is generally poor it is important to recognise that this measure is still generally effective in showing a high disruption value for the examples showing severe disruption and conversely a low figure for those with the least visually apparent disruption. This classification–evaluation

route does not seem to provide a sufficient efficiency in the detection of disruption to overcome the effects of other influences on the result, so that only the extreme cases are effectively charaterized.

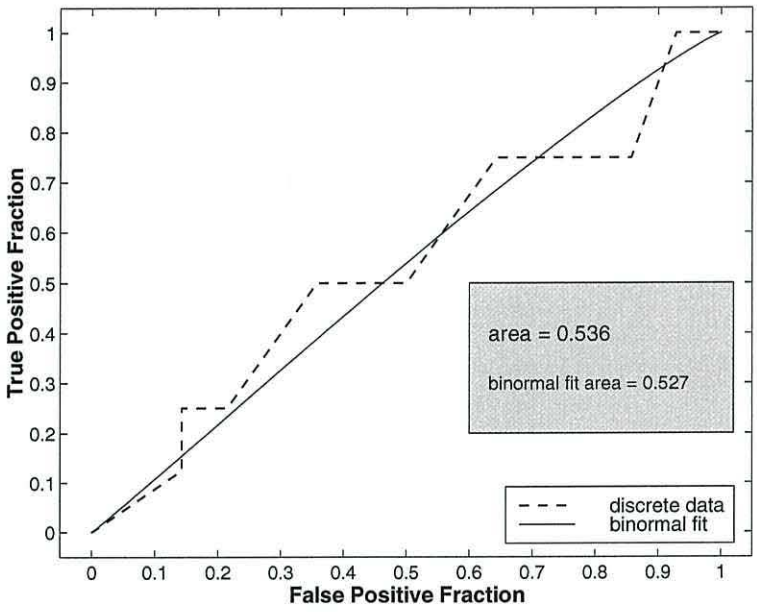


Figure 7.24: ROC curve and analysis for the final SOFM classification χ^2 metric results on the larger data set.

The area B evaluation was also performed on the SOFM classification and the results are similar to those shown in figures 7.23 and 7.24. The correlation coefficient is slightly higher ($\approx$0.56) as is the ROC area ($\approx$0.57) however these improvements are extremely small.

The generally poor performance of the SOFM classification method may relate to the way in which the network responds to similar profiles. The SOFM network is based on the idea that the provided input data instances (in this case the profiles) can be viewed as occupying a form of continuum with similar profiles being close together and distinct profiles further apart. Not only does SOFM require the number of classes to be pre-specified, but the network will use more of these classes in areas of the continuum which have a greater density of observations. This property has a number of consequences in terms of the analysis of skin pattern profile data as detailed below.

An image composed largely of undisturbed skin will result in large numbers of

similar profiles together with a few anomalies. The skin profiles then form a large density at one part of the continuum and the SOFM classification will respond to this by allocating a large number of classes for the representation of these essentially similar profiles. This then leaves relatively few classes which will model a large range of the anomalous profiles. The division of similar skin areas and the grouping together of infrequent profiles even where they are relatively different not only leads to classification images that are difficult to interpret visually, but also to other complications in terms of evaluation. Figure 7.25 shows this unwanted division of similar profiles for the test image of hand-drawn lines and a benign example: For the test image, the mRAC (centre) classification yeilds single classes for each consistent region, whereas the SOFM classification (right) always uses multiple classes. For the benign lesion example, the mRAC segmentation (left) shows most of the image in a single class, however the SOFM segmentation (right) is a confusion of classes.

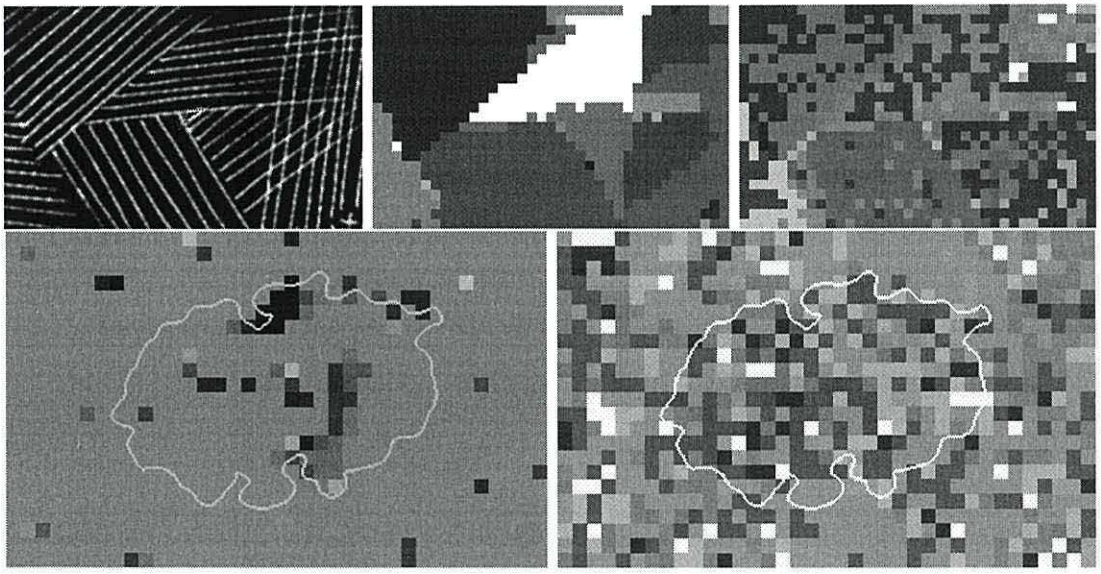


Figure 7.25: Top: Classifications for a test image contining several regions of consistent preferred orientations; From left to right the image, the mRAC classification and the SOFM classification. Bottom: Classifications for a benign lesion with large numbers of similar profiles. Left: mRAC classification showing a single large class. Right: SOFM classification, a confusion of many classes.

Division of essentially similar profiles into several different classes has significant conseqences in terms of evaluation through the proposed χ^2 or area B analyses. A small but consistent difference between the profiles as detected over the lesion

area compared to those of the skin area will cause the lesion are to be represented by completely different classes to the skin, in spite of the fact that the skin patterning and the profiles are essentially the same in both regions. The process of highlighting the skin lines for example, can result in the lesion area showing slightly more contrast than the surrounding skin (see figure 6.11 bottom left). Such a differentiation in classes is catastrophic in terms of χ^2 or area B analysis as the classification would then exhibit a significant correspondence to the skin and lesion areas resulting in a spurious high disruption estimate.

Another source of difficulty in terms of χ^2 and area B analysis is the lack of any consideration of locality; Each profile is considered as a separate entity to be classified in relation to its content alone. Since the classifier does not use locality information, classes can (and do) appear in many different places over the image. This class re-use means that the variety of profile types (and hence classes) in a disrupted lesion means that anomalous profiles in the skin area often fall into classes used over the lesion (where they reflect a similar portion of their respective areas) and that some profiles from the lesion area will fall into the classes found in the skin simply by chance. Both cases result in a false reduction in the χ^2 result for disruptive lesions only.

Although the χ^2 and area B analyses are not effective for the SOFM classification, disruption information still seems to appear in the classification images. Figure 7.26 shows the SOFM classification result for a malignant and a benign lesion example (left top and bottom respectively). The malignant lesion appears to use a different and much larger set of classes than the skin that surrounds it, whereas the benign example has no such correlation. These observations are however in many ways illusory; even in the particularly high disruption malignant case shown $\frac{3}{4}$ of the classes still appear in the skin area so that although every class appears in the lesion area, this represents neither a radically different nor considerably larger set of classes.

The observed differences between classifications for disrupted and non-disrupted skin patterning are actually a result of increased 'fragmentation' over the lesion area. This cannot of course be the same fragmentation as used in the mRAC evaluation for three main resons: Firstly, the overall number of classes is not

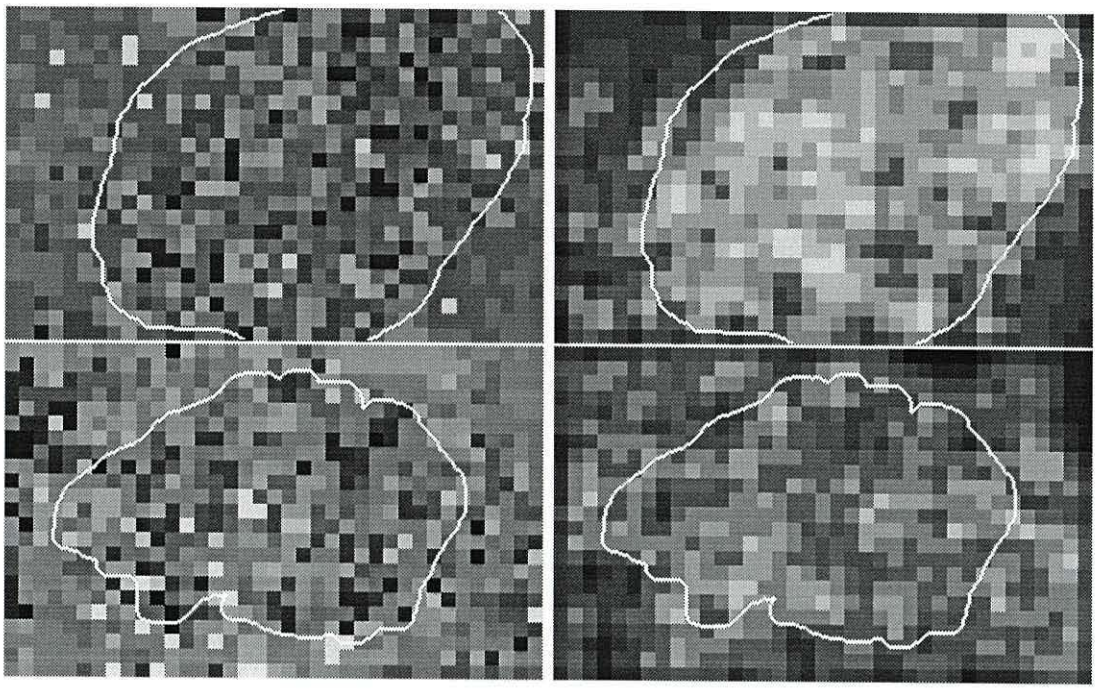


Figure 7.26: Local fragmentation results for a melanoma (top) and benign naevus (bottom). Left: SOFM segmentation result and right local fragmentation results (3x3 window class count).

free to vary so that the number of classes used for disruptive lesions will be falsely low and where the skin pattern is highly similar throughout the number will be falsely high. Secondly, the SOFM density matching property previously described will cause areas of essentially consistent patterning to use many classes and disrupted areas to use fewer. Finally, global class re-use means that an area of disrupted pattern contains much fewer classes than is immediately apparent; although neighbouring profiles rarely fall into the same class, a similar profile often arises in several different places in the lesion area.

The increased fragmentation observed in fact only reflects an increase in the number of classes used in a small local area. Figure 7.26 illustrates this property with an image showing the number of classes in a 3x3 window centred at each point for both of the SOFM classifications shown. The top row is for a melanoma and shows a clear increase in local fragmentation corresponding with the lesion extent, the bottom is for a benign naevus which shows little difference between the two regions.

The local class count information can be used as a disruption measure in much

the same way as the local variation measure, by taking an average value for the lesion and skin regions and looking at the difference. In this case, comparing the median value is not effective as the local class count is a discrete result allowing only a small range of values. Comparing the means provides a continuous result albeit still only with a small range. Using this measure of disruption an ROC area of $\approx$0.7 is obtained indicating that the SOFM classification is indeed capturing at least part of the desired information on disruption. The correlation to the visual estimate is however relatively poor, especially for the examples where the skin patterning is less well defined or detected.

The many problems associated with the SOFM classification, together with its overall poor performance suggest that this classification–evaluation route would not be suitable as a measure of patterning disruption.

7.5.4 ART Results

The fourth set of results are for the ART classification again using the CB-distance technique for the similarity criterion. The χ^2 results are shown in figure 7.27. The correspondence with the visual estimate is better than that obtained through the SOFM evaluation although the spread is still greater than with the local variation and mRAC fragmentation measures. The resuting correlation coefficient is $\approx$0.70. Performance with respect to the histological diagnosis is unimpressive in comparison with the local variation and mRAC fragmentation measures; No clear separation between the two lesion categories is achieved although the melanoma examples do show generally higher disruption figures than the naevi.

The ROC curve for this data clearly reflects the lack of a clear distinction between the results for the two lesion categories with an area of $\approx$0.61. The performance is clearly better than that for the SOFM evaluation and although the performance is still generally poor, in this case not only is a good result being obtained for the examples visually classified as extremes, but also the two categories are showing different characteristics in their results. The area B evaluation was also performed on the ART classification and the results are very similar to those shown in figure 7.23. The correlation coefficient is slightly lower ($\approx$0.66) but the ROC area is

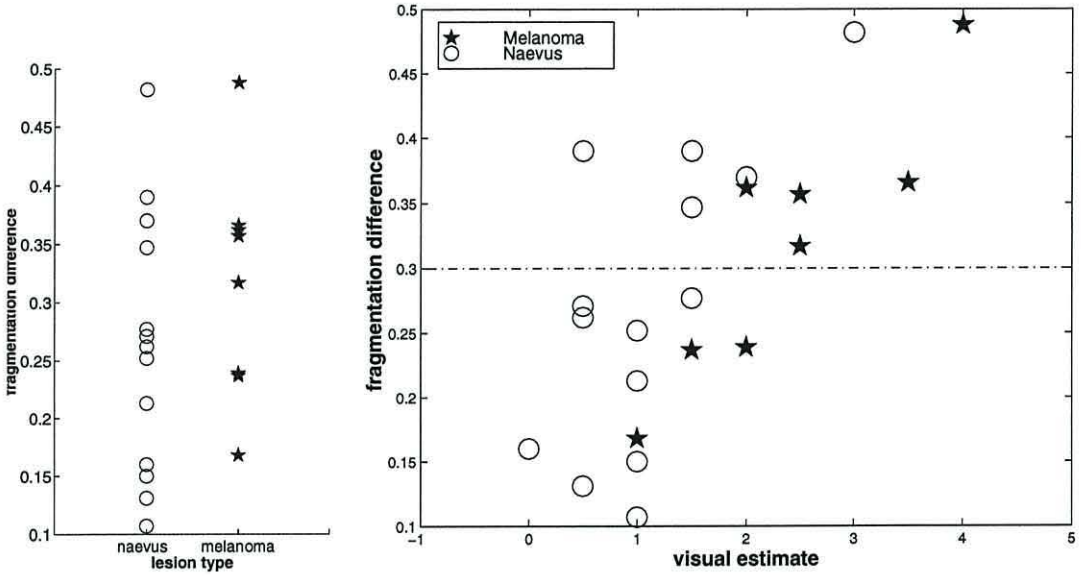


Figure 7.27: Final ART χ^2 evaluation results on the larger data set.

unchanged ($\approx$0.67). In all, this classification–evaluation route does not seem to provide a sufficient 'efficiency' in the detection of disruption to overcome the effects of other influences on the result.

The poor performance of the χ^2 and area B measures on the ART classification is perhaps due to the effects of class re-use as described above, where the classes used for a disrupted lesion are also found in the skin area where they are representing anomalous patterning and where skin classes are found in a disrupting lesion simply by chance. The ART network does not have the density matching property of SOFM and its consequent problems in terms of χ^2 or area B analysis and this may well explain the improvement in performance. The local class count evaluation is also applicable to the ART classification and the results are shown in figure 7.28 with the corresponding ROC curve in figure 7.29. These results clearly demonstrate the reduction in performance of the χ^2 and area B measures resulting from the absence of any consideration of locality.

The correspondence with the visual estimate is good although there is a significant range of results for those cases visually showing low degrees of patterning disruption. The correlation coefficient is $\approx$0.83 which is comparable to that obtained using either the local variation or mRAC fragmentation measures. Performance with respect to the histological diagnosis is also good, however although there is

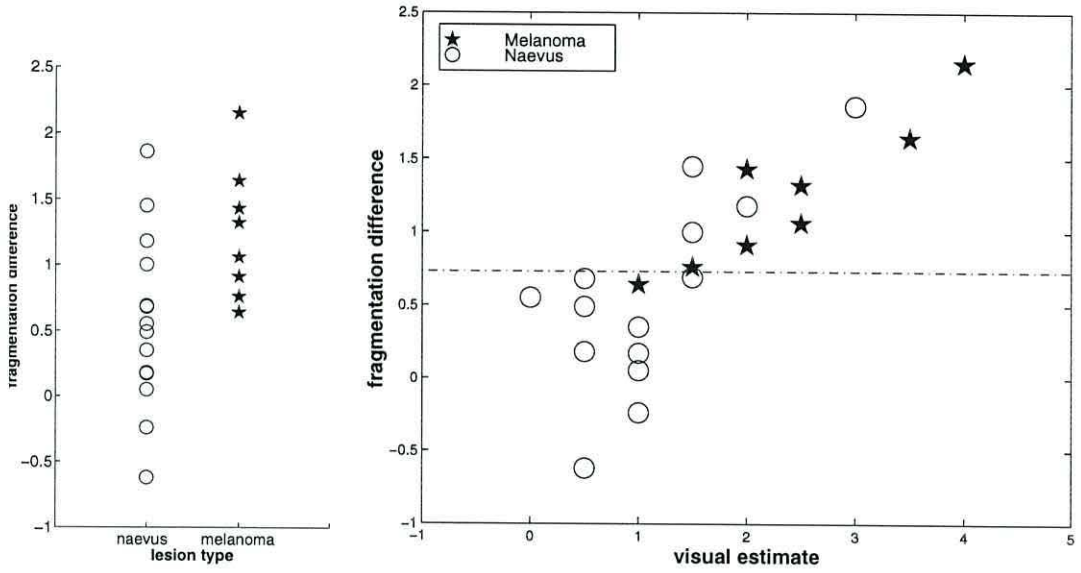


Figure 7.28: Final ART difference of mean local-class-count evaluation results on the larger data set.

a clear difference in the response for the two categories there is also considerable overlap between the results for the more disrupting naevi and the melanoma examples. The ROC curve for this evaluation reflects the good separation between the less disruptive naevi and the melanoma examples with an area of ≈0.80.

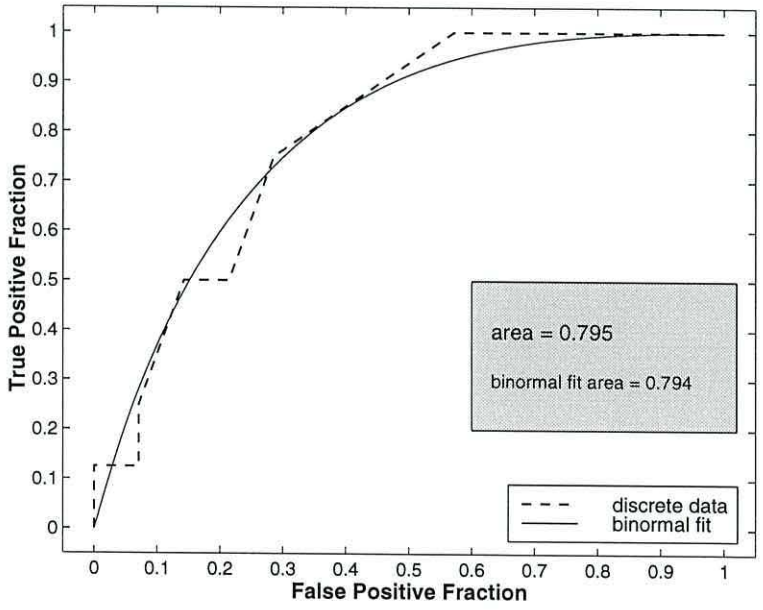


Figure 7.29: ROC curve and analysis for the final ART classification local-class-count metric results on the larger data set.

Although the local class count evaluation yields a reasonable discrimination performance it is not immediately clear what benefit this more complex route offers over the more direct local variation measure. The ART classification step adds more parameters and computation in addition to any effects of the peculiarities of the ART system without a clear improvement in performance. The use of a number of conceptually different and separate classification–evaluation routes is however desirable in order to improve robustness, so that given developments in other parts of the skin patterning acquisition system would warrant the inclusion of ART in testing of evaluation methods.

The four sets of results clearly indicate that the consideration of locality in a measure of disruption is essential. The consistency of normal skin pattern structure apparent to the human observer seems to relate to the small local regions having the same profile of preferred orientations and although to the human eye the patterning often remains similar over extended regions of skin, a non-region based classifier like SOFM or ART (when set to be sufficiently sensitive to separate the different classes due to a disruptive lesion) detects the presence of several different classes of profile which share this skin area. The region-based property of mRAC allows the use of a less stringent threshold so that these small differences in the skin are less influential whilst still maintaining separate classification for profiles with no similar neighbours as in regions of patterning disruption. The use of locality information is also clearly inherent in the local variation measure. However, this evaluation *only* considers the local information and so cannot account for the observed consistency of skin patterning over extended areas. The use of both immediate and extended regional consistency information as found in the mRAC fragmentation evaluation seems to be the key to successful analysis of skin patterning disruption.

7.6 Conclusions

Quantitative interpretation of the profile matrix is a major issue if this process is to be useful in an automated detection system. Methods based on local variability and on classification of the profile matrix obtained from mRAC and SOFM

revealed the differences expected between malignant and benign lesion examples. This information can be used together with the boundary detection methods to provide information about the disruption of the skin surface within the lesion. Initial trials based on the autocorreltation profiling method indicated that metrics for skin patterning disruption could be obtained which were useful in discriminating between melanoma and naevi lesions.

Detailed analysis of the profiling stage resulted in the development of a replacement for the autocorrelation method. The new CHP profiling method provided improvement in both sensitivity and fidelity to the underlying skin pattern. Simple enhancement and noise reduction measures were developed to improve the quality of representation and to account for variations in enhanced image quality (from effects such as loss of focus). The final threshold–stretch scheme is effective in removing noise and variability in profiles whilst preserving the detection of the skin patterning character.

The analysis and evaluation of the matrix of profiles relies on the notion of similarity. In terms of the skin patterning represented, the comparison of two profiles using standard vector distance measures such as L2 ignores important aspects of the information they embody. A new measure was developed to account for the toroidal nature of the profile vector and to reduce the effect of small differences in orientation. The new CB-distance measure was integrated into a replacement for the original variance based analysis of local variability. The new technique results in a more faithful representation of local pattern variability, avoids the blurring of the variance technique and allows a better explanation of exactly what this result means in terms of the skin patterning.

The analysis of a classification is an important part of the majority of the techniques investigated for the extraction of a metric of patterning disruption from a matrix of profiles. In order to facilitate the comparison of different classifiers and the assessment of new techniques, the possibility of a single parameterless evaluation for any classification was investigated. The area B evaluation was developed from the concepts employed in the individual evaluations used in the initial trials and its conceptual similarity to the standard statistical χ^2 measure was discussed. The problems relating to inherent assumptions concerning the

nature of the classification common to both these measures were analysed and the consequent unsuitability of both for the region-based mRAC classification was detailed. A simple replacement suitable for the mRAC classifications of the new enhanced CHP data was presented.

The ART classifier dismissed due to confusing classification results in the early trials, was re-tried on the enhanced CHP profile data since the adverse effects of the internal normalization would be reduced by the individual stretching of profiles performed in profile enhancement. The resulting classifications showed promise although the instability of the ART classification between runs on the same data was noted.

Final results for an expanded set of example melanoma and naevus images showed the local variation and mRAC fragmentation metrics performing well both in classification of the two lesion types and particularly in comparison to a visual estimate of the level of patterning disruption caused by the lesions. A particularly good separation of the two lesion diagnoses is obtained using the mRAC fragmentation evaluation. The results for the SOFM classification with both χ^2 and area B evaluations are poor. This failure can be traced to the unsuitability of the SOFM 'density matching' property and the problems of classification without consideration of locality information. A simple locality based evaluation which considers the consistency of classification in a small neighbourhood shows that some diagnostically useful information is still apparently available when the importance of locality is included in the analysis of a profile matrix. The ART classification also yields generally poor results with both χ^2 and area B evaluations, although some improvement over the SOFM results is apparent. Use of the locality measure employed on the SOFM classification results in a metric which performs comparably to the mRAC fragmentation and local variability measures, however the instability noted for ART together with the lack of clear advantage for this evaluation compared with the simpler and less computationally intensive local variation metric casts doubt on the inclusion of ART in further trials.

In all, the acquisition and modelling of skin line patterning from clinical images of skin lesions has been successfully achieved and the analysis of the resulting data has been shown to provide an assessment of pattern disruption which is both

consistent with visual inspection and effective in presenting information useful for discrimination between melanoma and benign naevi lesion examples.

Chapter 8

Conclusions

8.1 Overview

The research presented here has considered automatic skin cancer diagnosis. The role of computer-based diagnosis, and its value within a primary care situation were examined resulting in synthesis of aims, requirements and properties for an effective system — a system based on digital optical images captured and processed using low-cost, commercial computer technology.

The issues involved in acquisition of lesion boundaries were discussed. The value of accurate and robust boundaries, in terms of both directly obtainable diagnostic features and in enabling lesion property evaluation, was identified. Previous research proposed the edge focusing process. This work has addressed the improvement, in terms of potential for future development, evaluation and re-use, of this process through porting it to a highly modular form in the Khoros environment.

The role of colour analysis and its value in terms of provision of diagnostically useful features was investigated, and the central importance of segmentation identified. The fundamental properties of effective segmentation of lesion image colours were identified as a need to reflect human perception of colour similarity and a basis on local regions. A new region-based segmentation technique using data

transformed to a perception-uniform colour-space was presented and shown to yield promising results.

Finally the use of texture information was discussed. The nature and properties of the large-scale texture of skin patterning and its disruption were investigated and an abstracted representation proposed. A new technique was presented and shown to be effective in extracting the qualities of the skin patterning. Methods for analysing this representation of the patterning to quantify the disruption attributable to the lesion were proposed and developed. The combination of these extraction, analysis and disruption evaluation techniques was shown to be effective in relation to both visual assessment of disruption and diagnostic performance.

8.2 Detail

Background

The incidence of skin cancer is high and rising across the world, a trend which is likely to continue given the importance of solar UV as a cause and the modern propensity to indoor lifestyle with sun-seeking holidays. Malignant Melanoma is the most deadly form due to its rapid development, invasion and metastasis cycle. Early diagnosis is paramount as prognosis is greatly improved where the tumour is excised quickly. Many common benign pigmented skin lesions can resemble early melanomas which means that diagnosis *in situ* is both necessary, as it is not practical to excise all such lesions, and difficult as the early differential signs are hard to detect without detailed expert inspection.

System Concept Analysis

The task of providing diagnosis support to aid in the fight against skin cancers is far from simple. The discrimination task itself is widely recognized as difficult even for experts in the field, and more so for early lesions. The priorities of earlier diagnosis and quicker referral suggested a focus on improving the speed and accuracy of primary care diagnosis. The breadth of knowledge demanded of a GP precludes the specific depth required given the difficulty and highly skilled

nature of the discrimination task and this indicated the need for a diagnosis support tool.

An effective diagnosis support tool for the primary care sector must satisfy the key requirements of minimum capital cost, minimum training for operation, and minimum operating time, cost and fuss. These factors indicated an image based system avoiding specific requirements on image capture and using simple off-the-shelf and therefore low cost PC technology. The diagnosis needs to be highly sensitive so as to avoid missing any cancers whilst retaining a satisfactory specificity to avoid excessive referrals. The complex nature of the discrimination task suggested the development of a wide range of mutually independent feature estimates. These should be related to known medical indicators or visible features in order that confidence in the diagnosis can be boosted through interrogation of the "reasoning" underlying the final diagnosis.

Boundary Finding, Edge Focusing and Khoros

Robust and accurate identification of the extent of a lesion is a vital element of any image based skin lesion diagnosis support system. Boundary information directly enables the estimation of a variety of important diagnostic indicators such as size, shape, asymmetry and irregularity. In addition, boundary information enables the evaluation of features and properties both for the lesion area in isolation, and comparatively to the surrounding skin. Boundaries drawn by hand suffer from both inter and intra-observer inconsistency and are further hampered by the difficulty in providing even a good definition of exactly what constitutes the boundary of a lesion. Automatic boundary detection has consequently attracted considerable research interest. In previous research the edge focusing technique has been proposed and shown to be effective and reasonably robust as a means of accurate automatic lesion boundary detection.

The existing edge focusing system was reconstructed in the Khoros environment in order to facilitate the further development, improvement, evaluation and integration of the technique. The entire process has been decomposed into its functional elements and then rebuilt within the Cantata visual programming environment. Within Khoros, the implementation the edge focusing process is presented in a highly visual manner with the basic processing actions linked by data flow lines.

Evaluation in terms of the goals driving the conversion have shown the value of this presentation; it is ideal for experimentation since elements can be readily exchanged, allowing for simple evaluation of alternatives and for reuse of elements in future projects, and it greatly increases development prototyping and testing turnover by allowing access to all data passed between the process elements and by providing a variety of data visualization and analysis tools which can simply be 'plugged in' to examine the data at any point. It was noted that the re-implemented system had a significantly longer execution time due to the overhead involved in data flow using permanent files. However if speed were to become an important issue, Khoros provides the option of using much faster, non-permanent (but consequently non-interrogable) data transports.

In general the edge focusing technique is effective in providing robust and accurate lesion boundaries, however certain situations (such as lesions with highly indistinct boundaries and lesions which extend to the edge of the image) can cause poor results or failure. Solutions were proposed for some of the identified problems including the use of the Dull Razor technique for removing hair features from the image. The boundaries used in this study were governed by a boundary definition policy which is in essence that: boundaries were obtained, wherever possible, from the edge focusing system, but where the automatic system is ineffective or fails, boundaries were constructed by hand.

Colour Analysis and Segmentation

Colour information is widely recognized as an important diagnostic feature in the analysis of skin lesions, and this is reflected in its prominence in the skin cancer identification checklists. The colour-based indicators used for diagnosis mainly relate to either the variegation of colouring or the presence of specific colours. Research into colour-based lesion diagnosis falls into the three categories of spectral analysis of accurately calibrated reflectance spectra, analysis of spectral properties of image-based information, or the analysis of colour images. The aims and requirements of the envisaged diagnosis support system indicated an image based system using techniques of the third category, since special equipment is required by the first and the second demands not only accurate calibration, but often additional supplementary data.

Colour image processing for skin lesion images involves identification of either irregularity in pigmentation, specific colours, or the lesion extent. All these goals essentially reduce to segmentation of the image into regions of distinct colour. There has been considerable interest in colour-based segmentation for lesion images aimed at either finding a binary lesion/skin division, or detecting variegated colouring. The reported methods often begin by reducing the dimensionality of the colour data, the methods used (e.g. PCT_1) however have been shown to result in much of the *colour* information being lost. The underlying goals of segmentation for lesion image analysis were examined in detail and the importance of a clear definition of the terms in the phrase 'areas of uniform colour' was exposed. The majority of the segmentation methods which have been used on lesion images consider the pixels as individual entities ignoring their location in the image, and require the number of output 'colours' to be specified in advance. The former is contrary to the goal of identifying contiguous regions in that the spatial position of the pixels is not being considered, whilst the latter is obviously problematic since variegated colour is a known differential indicator.

The demand for consideration of spatial information prompted the development of a region-based approach to lesion image segmentation. The mRAC agglomerative clustering technique was developed and was shown to be promising in identification of homogeneous regions of colour using the full RGB data and the L2 distance measure for colour similarity. There were however identifiable discrepancies between the colour grouping given by L2 on RGB data and that of the human observer. Many colour-spaces have been used in image processing to satisfy the requirements of a wide variety of applications. Three colour-spaces were seen to be particularly relevant to the need for approximation to human perception: RGB, IHS and $L^*u^*v^*$. An investigation the general properties of, and the distribution of the colour information from lesion images in, these spaces clearly indicated the value of $L^*u^*v^*$ with its decoupling of luminosity and chromaticity information and its construction on the principle of consistency between L2 distance in the space and perceived colour difference.

Using mRAC with the $L^*u^*v^*$–L2 combination of colour-space and colour difference measure, a good segmentation was achieved on the majority of the sample images available for this study, with distinct identifiable improvements over the

previous results in all but one of the illustrative examples. The desire to ensure consistency in what is considered to be a 'different colour' suggested the use of a fixed parameter set, however this prevented the current implementation from producing a satisfactory result in some (apparently) dark, low contrast cases. The $L^*u^*v^*$–L2 combination was able able to produce a good segmentation for such difficult cases simply by changing a single parameter to make the merging process stop earlier i.e. be more 'fussy' in terms of colour similarity. This further demonstrated the value of the transformed colour-space as no such solution could be obtained using the RGB–L2 combination — separation of the apparent 'colours' in the lesion can only be maintained in a highly complicated and confused segmentation.

The promising results obtained with the new method suggest that the segmentation information should be used not only for the quantification of variegation and presence of lesion colours, but also to provide support in boundary identification. However, the development required to convert the multi-region segmentation to a binary lesion/skin division and for investigation of combination methods and confidence rating using multiple boundary estimates could not be addressed in the time available.

Texture — Skin Patterning

The pattern of fine lines criss-crossing the surface of normal skin forms a macro-scale texture. Changes in the skin patterning texture have been identified as indicative of early melanoma. A measure of disruption of skin line patterning would then clearly be of benefit. Existing research relating to skin patterning texture extends only to changes in properties of skin surface profiles taken perpendicular to the direction of the main skin lines. It is clear however that considerable potential for diagnostic information exists in the disruption of the directional pattern formed by the skin lines.

The nature of skin patterning means that it is poorly described by existing texture techniques and consequently a new skin patterning detection method has been developed. A matrix of profiles (by angle) of linear element strength is created, in effect a map for the preferred orientations of the patterning on the image. The process begins with a simple pre-processing method to highlight the skin lines in

the lesion images. Patches spaced evenly over the entire image are then analysed and each produces a profile representing the character of the local skin patterning. This process has been shown to be effective in relation to test patterns and in reflecting the observed skin patterning in real lesion images.

Quantitative interpretation of the profile matrix is a major issue if this process is to be useful in an automated detection system. Methods based on local variability and on classification of the profile matrix obtained from mRAC and SOFM provide patterning similarity information which can be combined with boundary data to yield a measure of the disruption of the skin surface within the lesion. Initial trials based on the autocorrelation profiling method showed promising results.

A detailed analysis of the profiling stage resulted in the development of a replacement for the autocorrelation method. The new CHP profiling method provided improvement in both sensitivity and fidelity to the underlying skin pattern. Simple enhancement and noise reduction measures were developed to improve the quality of representation and to account for variations in enhanced image quality (from effects such as loss of focus). The final threshold–stretch scheme was effective in removing noise and variability in profiles whilst preserving the detection of the skin patterning character.

The analysis and evaluation of the matrix of profiles relies on the notion of similarity. In terms of the skin patterning represented, the comparison of two profiles using standard vector distance measures such as L2 ignores important aspects of the information they embody. A new measure was developed to account for the toroidal nature of the profile vector and to reduce the effect of small differences in orientation. The new CB-distance measure was integrated into a replacement for the original variance based analysis of local variability. The new technique results in a more faithful representation of local pattern variability, avoids the blurring of the variance technique and allows a better explanation of exactly what this result means in terms of the skin patterning.

The analysis of a classification is an important part of the majority of the techniques investigated for the extraction of a metric of patterning disruption from a matrix of profiles. In order to facilitate the comparison of different classifiers and the assessment of new techniques, the possibility of a single parameterless

evaluation for any classification was investigated. The area B evaluation was developed from the concepts employed in the individual evaluations used in the initial trials and its conceptual similarity to the standard statistical χ^2 measure was discussed. The problems relating to inherent assumptions concerning the nature of the classification common to both these measures were analysed and the consequent unsuitability of both for the region-based mRAC classification was detailed. A simple replacement suitable for the mRAC classifications of the new enhanced CHP data was presented.

The ART classifier dismissed due to confusing classification results in the early trials, was re-tried on the enhanced CHP profile data since the adverse effects of the internal normalization would be reduced by the individual stretching of profiles performed in profile enhancement. The resulting classifications showed promise although the instability of the ART classification between runs on the same data was noted.

Final results for an expanded set of example melanoma and naevus images showed the local variation and mRAC fragmentation metrics performing well both in classification of the two lesion types and particularly in comparison to a visual estimate of the level of patterning disruption caused by the lesions. A particularly good separation of the two lesion diagnoses was obtained using the mRAC fragmentation evaluation. The results for the SOFM classification with both χ^2 and area B evaluations were poor. This failure can be traced to the unsuitability of the SOFM 'density matching' property and the problems of classification without consideration of locality information. A simple locality based evaluation which considers the consistency of classification in a small neighbourhood shows that some diagnostically useful information is still apparently available when the importance of locality is included in the analysis of a profile matrix. The ART classification also yielded generally poor results with both χ^2 and area B evaluations, although some improvement over the SOFM results was apparent. Use of the locality measure employed on the SOFM classification resulted in a metric which performs comparably to the mRAC fragmentation and local variability measures, however the instability noted for ART together with the lack of clear advantage for this evaluation compared with the simpler and less computationally intensive local variation metric casts doubt on the inclusion of ART in further

trials.

In all, the acquisition and modelling of skin line patterning from clinical images of skin lesions has been successfully achieved and the analysis of the resulting data has been shown to provide an assessment of pattern disruption which is both consistent with visual inspection and effective in presenting information useful for discrimination between melanoma and benign naevi lesion examples.

8.3 Further Work

This work has highlighted a number of key areas for future consideration in relation to the construction of the diagnosis support system in general, and for future research in relation to the specific colour and texture techniques presented.

General system issues.

Problems with the initial image have been seen to have an impact on boundary finding, colour and texture processing. Simple controls on the initial image capture would be the ideal solution for many of these problems; the adoption of a fixed, but wider field of view and the addition of colour and intensity matching charts into the image would ensure a reasonable area surrounding the lesion and avoid generally dark and low contrast images.

A standard sequence of processing to improve the quality of the captured image should also be considered. The impact of non-uniform illumination and variations in focus could be reduced with illumination surface correction and re-focusing techniques, and additional image processing to remove 'noise features' would also be valuable. It is however important to ensure that quality improvement processing preserves all the required data in the image; the DullRazor hair removing technique for example, employs smoothing and hence results loss of fine detail in parts of the image. Careful construction of an effective but 'safe' image improvement sequence including a modified DullRazor would however be invaluable, and research and development with this aim would form a vital part of the diagnosis support system programme.

Performance evaluation is a complex issue in lesion image analysis. Evaluation of the techniques presented in this study has often been in relation to human visual assessment of the feature in question. This form of assessment is essential in view of diagnosis support system requirements for visually verifiable features and explainable diagnosis, in relation to the essential properties of good colour segmentation, and especially in ensuring that the intended feature is indeed being assessed by the technique as with the skin patterning disruption measure. However such assessment is problematic in terms of comparative performance evaluations due to its inherent subjectivity. Quantitative feature evaluations are however hampered by the need for a 'gold standard' both in feature definition and quantification, i.e. the need to know the correct response for the given feature. Furthermore, comparative performance evaluation for different processes in terms of features requires a standard set of images with accompanying feature data for each image to be generally available. Without a fixed test set performance differences could simply be the result of differing imaging processes, feature definitions or feature assessment. Evaluation in relation to lesion type given by histological diagnosis is common and avoids the feature definition and assessment problems, however features generally do not have a *perfect* correspondence to diagnosis (clearly shown by the overall score from a list of features being used to signify diagnosis in published checklists) so that diagnostic performance does not necessarily quantify performance concerning the feature in question. An image library with associated 'feature files' has been used in several studies by Umbaugh *et al* e.g. [72] and Lee *et al* [58] use a set of lesion images compiled with scored feature data. Neither of the image sets is generally available, the latter cannot be released for legal reasons. A more general dermatology image atlas, DOIA [93] offers open access but only has general feature data and uses JPEG image compression which results in corruption of fine detail features such as skin patterning. Other lesion specific image libraries which may provide a solution are being established by the American Academy of Dermatology and by the UK Melanoma Study Group [33]. A full investigation of performance evaluation issues, and feature definition and assessment in test image sets would again be a vital part of the diagnosis support system programme.

Colour Analysis.

The identified importance of locality information for colour segmentation of lesion

images was addressed in this work through the development of the mRAC region-based process. An alternative could exist in the encoding of the location as part of each pixel value. This would allow limited locality consideration within the many segmentation algorithms which consider each pixel as separate entities (those which do not require pre-specification of the number of classes would be most suitable). Simple location encoding may not be successful as it would normally introduce a bias toward circular regions (where each pixel is 'close' to the mean location of the class), investigation of more complex encoding including the notion of connectivity as part of 'nearness' could yield a more successful result. Such trials would be of greatest importance if much higher resolution images need to be segmented as the iterative nature of the mRAC process makes it relatively computationally intensive.

Most colour segmentation schemes begin with smoothing. The mRAC process does not require such pre-processing with the consequent advantage of avoiding the associated blurring of edges. The application of smoothing may however allow a greater sensitivity to colour difference by reducing the relative noise level and hence allowing the merge phase to stop earlier. Any smoothing applied should be sensitive to the blurring problem, small window iterative median (or less computationally expensive psuedo-median) filters suppress noise with relatively little blurring impact [37], however adaptive shape filtering by taking the median of the least variable contiguous segment of the filter window may be more effective. Filtering applied to a higher resolution version of the image before sub-sampling could also be a valuable approach.

The assessment of colour similarity formed an important element of this study. Consideration of the distribution of image data in the colour-space allows the investigation of the correspondence between areas of perceptually similar colour and groupings in the colour-space and consequently the identification of suitable colour similarity measures. Such analysis could be expanded from the selection of a suitable colour-space and distance measure pair, to the formulation of a dedicated colour difference measure specifically tailored to the needs of lesion analysis, for example differences in blue-black shades may be more important than differences in shades of pink. The concept of colour-texture could also provide a means of better distinguishing between perceived colours, since skin

colour, for example, is obviously a perceptual combination of different colours.

The colour segmentation information should be used not only for the quantification of variegation and presence of lesion colours, but also to provide support in boundary identification. Conversion of the multi-region segmentation to a binary lesion/skin division and investigation of combination methods and confidence rating using multiple boundary estimates should be addressed. Texture data may also provide new boundary data, a multi-channel approach for the estimation of a feature provides increased robustness and a means to quantify of confidence a feature measure. A full investigation of the issues and techniques, and an awareness of the value of cross-channel support would form an important part of the programme as a whole.

Texture Analysis

The skin patterning texture analysis begins with highlighting of the skin lines performed by a smooth model subtraction method. Although this has been shown to provide an adequate representation, a refinement of this stage may yield increased robustness to poor image quality. In terms of initial pattern highlighting, since small scale texture models must consider skin patterning as noise, subtraction of the texture generated from such a model ought perhaps to be effective in revealing the patterning, however in practice this method may be ineffective as the generated texture might not be well registered (i.e. synchronized) with that of the image. The FFT based processing should also be re-considered when faster hardware is available. Cleaning of the highlighted pattern is perhaps the most likely area for improvement. An adaptation of a morphological skeletonization technique may be successful and trials of the logical/linear operator technique proposed by Iverson *et al* [94], in view of the impressive results they present for fingerprint line tracing with their L/L positive contrast operator, should be undertaken.

An increase in robustness may also be possible by modifying the profile enhancement threshold-stretch process to use an adaptive noise threshold in proportion to the range of the raw data. This would allow more consistent patterning profile detection for areas of low contrast in the pattern highlighted image whilst maintaining the profile shape properties in sharp defined areas, based on the as-

sumption that the noise content of low contrast areas (often the result of loss of focus in the original image) is smaller than that of high contrast areas.

Finally in the evaluation stage, the impact of introducing a 'transition zone' at the lesion border which is viewed as neither lesion nor skin, should be assessed. This would improve tolerance to a poorly localised boundary or to situations where the extent of the impact of the lesion on skin patterning does not precisely match that of the pigmentation (assuming boundary identification is based on pigmentation extent as in this study). The transition zone must remain small however to avoid unnecessary discarding of data, especially where the lesion itself is small.

References

[1] D. J. Gawkrodger, *Dermatology.* Churchill Livingstone, 1992.

[2] D. J. Leffell and D. E. Bash, "Sunlight and skin cancer," *Scientific American*, vol. 275, pp. 38–43, July 1996.

[3] "Patient information: Skin cancer updates."
American Academy of Dermatologists
http://www.aad.org/.
version current at 26 June 1998.

[4] "Sun and skin: a dangerous combination."
Australian Academy of Science
http://www.science.org.au/nova/008/008key.htm.
version current at 26 June 1998.

[5] R. Marks, *Roxburgh's Common Skin Diseases.* Chapman and Hall Medical, 16th ed., 1993.

[6] USA National Institutes of Health, "Diagnosis and treatment of early melanoma," *NIH Consensus Statement*, vol. 10, pp. 1–26, Jan. 1992. document also available online:
http://text.nlm.nih.gov/nih/cdc/www/88txt.html.

[7] R. M. MacKie, "An illustrated guide to the recognition of early malignant melanoma." printed by Blackwood Pillans & Wilson Ltd. Edinburgh, Department of Dermatology, University of Glasgow, 1986.

[8] P. N. Hall, "Clinical diagnosis of melanoma," in *Diagnosis and Management of Melanoma in Clinical Practice* (N. Kirkham, D. W. K. Cotton, R. C.

Lallemand, J. E. White, and R. D. Rosin, eds.), ch. 4, pp. 35–52, New York: Springer-Verlag, 1992.

[9] R. M. MacKie, "Melanoma prevention and early detection," *British Medical Bulletin*, vol. 51, no. 3, pp. 570–583, 1995.

[10] R. Marks, "Prevention and control of melanoma: The public health approach," *CA Cancer Journal for Clinicians*, vol. 46, pp. 199–216, 1996.

[11] E. Claridge, P. N. Hall, M. Keefe, and J. P. Allen, "Shape analysis for classification of malignant melanoma," *Journal of Biomedical Engineering*, vol. 14, pp. 229–234, May 1992.

[12] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "An automatic color segmentation algorithm with application to identification of skin tumour borders," *Computerized Medical Imaging and Graphics*, vol. 16, no. 3, pp. 227–235, 1992.

[13] F. Ercal, M. Moganti, W. V. Stoecker, and R. H. Moss, "Detection of skin tumour boundaries in color images," *IEEE Trans. on Medical Imaging*, vol. 12, no. 3, pp. 624–627, 1993.

[14] A. P. Dhawan and A. Sicsu, "Segmentation of images of skin lesions using color and texture information of surface pigmentation," *Computerized Medical Imaging and Graphics*, vol. 116, no. 3, pp. 163–177, 1992.

[15] J. E. Golston, V. W. Stoecker, R. H. Moss, and I. P. S. Dhillon, "Automatic detection of irregular borders in melanoma and other skin tumours," *Pattern Recognition*, vol. 23, no. 11, pp. 1235–1247, 1990.

[16] W. E. Denton, *Boundary and Shape Recognition for Automated Skin Tumour Diagnosis.* PhD thesis, University of Wales: Bangor, 1998.

[17] W. E. Denton, A. W. G. Duller, and P. J. Fish, "Boundary detection for skin lesions: An edge focussing algorithm," in *Proc. 5th Int. Conf. on Image Processing and Applications, Edinburgh*, pp. 399–403, IEE, July 1995.

[18] W. E. Denton, A. W. G. Duller, and P. J. Fish, "Synthesis of skin images for image processing validation," in *Proc. 9th Int. Conf. on Biomedical Engineering*, pp. 445–447, Singapore, Dec. 1997.

[19] E. E. Bondi, D. E. Elder, G. Dupont, and W. H. Clark, "Skin markings in malignant melanoma," *Journal of the American Medical Association*, vol. 250, pp. 503–505, 1983.

[20] B. J. Connemann, H. Busche, J. Kreusch, H. Teichert, and H. H. Wolff, "Quantitative surface topography as a tool in the differential diagnosis between melanoma and naevus," *Skin Research and Technology*, vol. 1, pp. 180–186, 1995.

[21] A. D. Potorac, I. Toma, and J. Mignot, "In vivo skin relief measurement using a new optical profilometer," *Skin Research and Technology*, vol. 2, pp. 64–69, 1996.

[22] Rasure and Kubica, "The khoros application development environment," in *Experimental Environments for Computer Vision and Image Processing* (H. I. Christensen and J. L. Crowley, eds.), World Scientific, 1994.

[23] A. J. Round, P. J. Fish, and A. W. G. Duller, "Khoros implementation of edge focusing system for skin lesion analysis," in *Proc. 18th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 1187–1189 (4.5.4–7), 1996.

[24] G. Vines, "Get under your skin," *Inside Science No.78 (in New Scientist Vol.145 No.1960)*, Jan. 1995.

[25] "Melanoma and skin cancer information."
Melanoma and Skin Cancer Research Institute, The Melanoma Foundation and The Sydney Melanoma Unit (Australia)
http://www.med.usyd.edu.au/medicine/melanoma/.
version current at 26 June 1998.

[26] "Tumours of the skin."
'Matrix' Dematology Resources
http://matrix.ucdavis.edu/tumors.html.
version current at 26 June 1998.

[27] D. S. Rigel, "Malignant melanoma: Perspectives on incidence and its effects on awareness, diagnosis, and treatment," *CA Cancer Journal for Clinicians*, vol. 46, pp. 195–198, 1996. (Guest Editorial).

[28] "Info for health professionals: Actinic keratosis (ak)."
Skin & Cancer Foundation of Australia
http://www.scfa.edu.au/
Particularly Info for Health Professionals,
specifically http://www.scfa.edu.au/ak.html.
version current at 12 Aug 1998.

[29] D. A. Perednia, "What dermatologists should know about digital imaging," *Journal of the American Academy of Dermatology*, vol. 25, pp. 89–108, July 1991.

[30] V. P. Wallace, J. C. Bamber, D. C. Crawford, and P. Mortimer, "Spectroscopy for the assessment of pigmented skin lesions," in *Biomedical Optical Spectroscopy and Diagnostics*, (Orlando, Florida, USA), pp. 67–69, Optical Society of America (Washington DC), 1996.

[31] T. M. Johnson, J. W. Smith II, B. R. Nelson, and A. Chang, "Current therapy for cutaneous melanoma," *Journal of the American Academy of Dermatology*, vol. 32, May 1995.

[32] R. M. MacKIE, "Clinical recognition of early invasive melanoma: Looking for changes in size, shape and colour is successful," *British Medical Journal*, vol. 301, pp. 1005–1006, 1990.

[33] P. N. Hall, E. Claridge, and J. D. Morris-Smith, "Computer screening for early detection of melanoma — is there a future?," *British Journal of Dermatology*, vol. 132, pp. 325–338, 1995.

[34] A. Breslow, "Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma," *Annals of Surgery*, vol. 172, pp. 903–908, Nov 1970.

[35] F. Ercal, A. Chawla, W. V. Stoecker, H.-C. Lee, and R. H. Moss, "Neural network diagnosis of malignant melanoma from colour images," *IEEE Trans. on Biomedical Engineering*, vol. 41, pp. 411–416, Sept. 1994.

[36] A. Kjoelen, M. J. Thompson, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Performance of ai methods in detecting melanoma," *IEEE*

Engineering in Medicine and Biology Magazine, vol. 14, pp. 411–416, July/August 1995.

[37] G. A. Hance, S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Unsupervised color image segmentation: With application to skin tumor borders," *IEEE Engineering in Medicine and Biology Magazine*, vol. 15, pp. 104–111, Jan/Feb 1996.

[38] S. Cotton and E. Claridge, "Developing a predictive model of human skin colouring," in *Proc. SPIE*, vol. 2708, Medical Imaging 1996: Physics of Medical Imaging, pp. 814–825, Apr. 1996.

[39] S. E. Umbaugh, R. H. Moss, and W. V. Stoecker, "Automatic color segmentation of images with application to detection of variegated coloring in skin tumours," *IEEE Engineering in Medicine and Biology Magazine*, vol. 8, pp. 43–52, Dec. 1989.

[40] A. D. Vanker and W. V. Stoecker, "AI/DERM: Diagnosis of skin tumours," in *Proceedings. AAMSI Congress 84*, 1984.

[41] D. A. Perednia, J. A. Gaines, and T. W. Butruille, "Comparison of the clinical informativeness of photographs and digital imaging media with multiple-choice receiver operating characteristic analysis," *Archives of Dermatology*, vol. 131, pp. 292–297, Mar. 1995.

[42] M. Binder, M. Puespoeck-Schwarz, A. Steiner, H. Kittler, M. Muellner, K. Wolff, and H. Pehamberger, "Epiluminescence microscopy of small pigmented skin lesions: Short-term formal training improves the diagnostic performance of dermatologists," *Journal of the American Academy of Dermatology*, vol. 36, pp. 197–202, Feb. 1997.

[43] "CSIRO media release: Australian advance in skin cancer detection."
On line publication of a media release by CSIRO — Australia's Commonwealth Scientific and Industrial Research Organisation.
http://www.csiro.au/ communication/mediarel/mr1997/mr97114.htm.
version current at 26 June 1998.

[44] "Derma Instruments — MoleMax II."
Derma Instruments: on-line presentation of MoleMax II.

http://www.derma.co.at/.
version current at 13 August 1998.

[45] "TransLite — Nevoscope."
TransLite Inc.: on-line presentation of the Nevoscope.
http://www.tlite.com/.
version current at 13 August 1998.

[46] A. P. Dhawan and P. Kini, "Pigment pattern and depth of invasion analysis of skin lesions for malignant melanoma detection," in *Proc. SPIE*, vol. 1231, Medical Imaging IV: Image Formation, pp. 75–83, 1990.

[47] W. E. Denton, A. W. G. Duller, and P. J. Fish, "Robust boundary detection for skin lesions," in *Proc. 17th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 407–408, 1996.

[48] N. Cascinelli, M. Ferrario, T. Tonelli, and E. Leo, "A possible new tool for clinical diagnosis of melanoma: The computer," *Journal of the American Academy of Dermatology*, vol. 16, pp. 361–367, Feb. 1987.

[49] R. White, D. S. Rigel, and R. J. Friedman, "Computer applications in the diagnosis and prognosis of malignant melanoma," *Dermatologic Clinics*, vol. 9, pp. 695–703, Oct. 1991.

[50] D. A. Perednia, R. G. White, and R. A. Schowengerdt, "Localization of cutaneous lesions in digital images," *Computers and Biomedical Research*, vol. 22, pp. 374–392, 1989.

[51] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," in *Proc. 1st International Conference on Computer Vision*, pp. 259–268, 1987.

[52] J. Anthony Yezzi, S. Kichenassamy, A. Kumar, P. Oliver, and A. Tannenbaum, "A geometric snake model for segmentation of medical imagery," *IEEE Trans. Medical Imaging*, vol. 16, pp. 199–209, Apr. 1997.

[53] I. Mikić, S. Krucinski, and J. D. Thomas, "Segmentation and tracking in echocardiographic sequences: Active contours guided by optical flow estimates," *IEEE Trans. Medical Imaging*, vol. 17, pp. 274–284, Apr. 1998.

[54] F. Bergholm, "Edge focusing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, pp. 726–741, Nov. 1987.

[55] J. Kittler, J. Illingworth, J. Foglein, and K. Paler, "An automatic thresholding algorithm and its performance," in *IEEE Proc. Seventh Int. Conf. on Pattern Recognition*, vol. 1, pp. 287–289, 1984.

[56] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[57] Rasure and Young, "An open environment for image processing software development," in *SPIE/IS&T Symposium on Electronic Imaging*, vol. 1659, 14 Feb. 1992.

[58] T. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "DullRazor: A software approach to hair removal from images," *Comput. Biol. Med.*, vol. 27, no. 6, pp. 533–543, 1997.

[59] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Addison-Wesley, 1993. (reprinted with corrections).

[60] W. K. Pratt, *Digital Image Processing.* Wiley-Interscience, second ed., 1991.

[61] C.I.E., *Colorimetry.* Publication No. 15.2, A-1033 Vienna, P.O. Box 169, Austria: Central Bureau of the C.I.E., Commission Internationale de l'Eclairage, 2nd ed., 1986.

[62] G. Wyszecki and W. S. Stiles, *Colour Science.* New York: Wiley, 1957.

[63] D. Bourgin, "Color spaces faq."
Available from several places on the web, the version used comes from:
http://www.ach.uams.edu/~adh/FAQS/color-space.FAQ
Another (apparently identical content) version with improved html formatting can be found at:
http://mpa.cat.cbpf.br/color_faq.html
A different FAQ maintained by Charles A. Poynton with similar content can be found at:
ftp://rtfm.mit.edu/ pub/usenet/news.answers/graphics/colorspace-faq.
version current at 23 July 1998 (dated as "Date: 5/3/1996 Last update: 18/4/1995".

[64] "Fully utilizing Photo-CD images: article no. 4 – PhotoYCC color encoding and compression schemes."
From Kodak (digital science) Photo-CD technical papers archive under PhotoYCC Color Encoding and Compression Schemes.
http://www.kodak.com/ country/US/en/digital/techInfo/pcd-045.shtml.
version current at 23 July 1998.

[65] M. L. Zoler, "Pigment clues guide melanoma risk diagnosis," *Skin & Allergy News*, vol. 28, p. 22, Sept. 1997.

[66] W. Kuznar, "Dermatoscopy transforms invisible to visible in pigmented lesions," *Dermatology Times*, Nov. 1997.

[67] "Image analysis project: Skin lesions classifications."
from the Image Analysis Group of CSIRO Mathematical and Information Sciences (CMIS) a part of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO).
http://www.dms.csiro.au/~talbot/ee02/melanoma.html.
version current at 26 June 1998.

[68] R. Marchesini. *et al*, "In vivo spectrophotometric evaluation of neoplastic and non-neoplastic skin pigmented lesions-i. reflectance measurements," *Photochemistry & Photobiology*, vol. 53, no. 1, pp. 77–84, 1991.

[69] R. Marchesini. *et al*, "In vivo spectrophotometric evaluation of neoplastic and non-neoplastic skin pigmented lesions-ii. discriminate analysis between naevus and melanoma," *Photochemistry & Photobiology*, vol. 55, no. 4, pp. 515–522, 1992.

[70] S. Cotton, P. Hall, J. RashBass, and E. Claridge, "Assisting diagnosis of melanoma through the 'noninvasive biopsy' of skin lesions," in *Proc. Medical Image Understanding and Analysis*, pp. 177–180, University of Oxford, July 1997.

[71] S. E. Umbaugh, Y. Wei, and M. Zuke, "Feature extraction in image analysis: A program for facilitating data reduction in medical image classification," *IEEE Engineering in Medicine and Biology Magazine*, vol. 16, pp. 62–73, July 1997.

[72] S. E. Umbaugh, R. H. Moss, W. V. Stoecker, and G. A. Hance, "Automatic color segmentation algorithms with application to skin tumour feature identification," *IEEE Engineering in Medicine and Biology Magazine*, vol. 12, pp. 75–82, July 1993.

[73] Y. W. Lim and S. U. Lee, "On the colour image segmentation algorithm based in the thresholding and the fuzzy c-means techniques," *Pattern Recognition*, vol. 23, no. 9, pp. 935–952, 1990.

[74] D. K. Panjwani and G. Healey, "Markov random field models for unsupervised segmentation of textured color images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 939–954, Oct. 1995.

[75] C. M. Bishop, *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[76] B. D. Ripley, *Pattern Recognition and Neural Networks.* Cambridge: Cambridge University Press, 1996.

[77] A. J. Round, A. W. G. Duller, and P. J. Fish, "Colour segmentation for lesion classification," in *Proc. 19th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, pp. 582–585 (2.3.4–3), Nov. 1997.

[78] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Addison-Wesley, 1992.

[79] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1. New York: Addison Wesley, 1992.

[80] F. Liu and R. W. Picard, "Periodicity, directionality and randomness: Wold features for image modeling and retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, pp. 722–733, July 1996.

[81] R. Jain, R. Kasturi, and B. G. Schunck, *Machine Vision.* New York: McGraw Hill, Inc., 1995.

[82] K. I. Laws, "Textured image segmentation," Tech. Rep. 940, University of Southern California, Image Processing Institute, USCIPI, Jan. 1980.

[83] B. B. Chaudhuri and N. Sarkar, "Texture segmentation using fractal dimension," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 72–77, Jan. 1995.

[84] T. C. Parr, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "Statistical modelling of oriented line patterns in mammograms," in *Proc. SPIE*, vol. 3034 of *Medical Imaging 1997*, SPIE, Feb. 1997.

[85] T. C. Parr, R. Zwiggelaar, C. J. Taylor, S. M. Astley, and C. R. M. Boggis, "The detection of stellate lesions in digital mammography," in *Medical Image Understanding and Analysis 1997*, July 1997.

[86] A. J. Round, A. W. G. Duller, and P. J. Fish, "Lesion classification using skin line texture," in *Proc. Medical Image Understanding and Analysis*, pp. 169–172, University of Oxford, July 1997.

[87] A. J. Round, A. W. G. Duller, and P. J. Fish, "Skin patterning analysis for lesion classification and segmentation," in *Proc. 9th Int. Conf. on Biomedical Engineering*, pp. 448–450, Singapore, Dec. 1997.

[88] G. Carpenter, S. Grossberg, and D. Rosen, "ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition," *Neural Networks*, vol. 4, no. 4, pp. 493–504, 1991.

[89] T. Kohonen, *Self-organisation and Associative Memory.* Springer-Verlag, 1989.

[90] A. J. Round, A. W. G. Duller, and P. J. Fish, "A comparison of skin patterning feature analysis methods for lesion classification," in *Proc. SPIE* (K. M. Hanson, ed.), vol. 3338, Medical Imaging 1998: Image Processing (part 1), pp. 202–210, Feb. 1998.

[91] J. A. Swets, "Measuring the accuracy of diagnostic systems," *Science*, vol. 240, pp. 1285–1293, 1988.

[92] C. E. Metz, "Basic principles of roc analysis," *Seminars in Neuclear Medicine*, vol. 8, pp. 283–298, 1978.

[93] "'DOIA' Dermatology Online Image Atlas."
Provided as part of the dermIS project, University of Erlangen, Germany.

http://www.derma.med.uni-erlangen.de/ bilddb/index_e.htm.
version current at 15 August 1998.

[94] L. A. Iverson and S. W. Zucker, "Logical/linear operators for image curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 982–996, Oct. 1995.

[95] C. A. Morton and R. M. MacKie, "Clinical accuracy of diagnosis of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 138, pp. 283–287, 1998.