UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
TESI DI DOTTORATO

(SEMI)-AUTOMATED ANALYSIS OF MELANOCYTIC LESIONS

SUPERVISORE: Prof. Enoch Peserico

DOTTORANDO: Francesco Peruch

Contents

To my wife,
Valentina,
my constant source
of inspiration and encouragement

And to my parents,
Danillo and Giovanna,
for their love and sacrifices

Abstract

Melanoma is a very aggressive form of skin cancer whose incidence has constantly grown in the last 50 years. To increase the survival rate, an early diagnosis followed by a prompt excision is crucial and requires an accurate and periodic analysis of the patient's melanocytic lesions. We have developed an hardware and software solution named Mole Mapper to assist the dermatologists during the diagnostic process. The goal is to increase the accuracy of the diagnosis, accelerating the entire process at the same time. This is achieved through an automated analysis of the dermatoscopic images which computes and highlights the proper information to the dermatologist. In this thesis we present the 3 main algorithms that have been implemented into the Mole Mapper:

A robust segmentation of the melanocytic lesion, which is the starting point for any other image processing algorithm and which allows the extraction of useful information about the lesion's shape and size. It outperforms the speed and quality of other state-of-the-art methods, with a precision that meets a Senior Dermatologist's standard and an execution time that allows for real-time video processing;

A virtual shaving algorithm, which increases the precision and robustness of the other computer vision algorithms and provides the dermatologist with a hair-free image to be used during the evaluation process. It matches the quality of state-of-the-art methods but requires only a fraction of the computational time, allowing for computation on a mobile device in a time-frame compatible with an interactive GUI;

A registration algorithm through which to study the evolution of the lesion over time, highlighting any unexpected anomalies and variations. Since a standard approach to this problem has not yet been proposed, we define the scope and constraints of the problem; we analyze the results and issues of standard registration techniques; and finally, we propose an algorithm with a speed compatible with Mole Mapper's constraints and with an accuracy comparable to the registration performed by a human operator.

Sommario

Il Melanoma è una forma molto aggressiva di cancro alla pelle la cui incidenza è costantemente aumentata negli ultimi 50 anni. Una diagnosi precoce unita ad una rapida asportazione risulta indispensabile per migliorare il tasso di sopravvivenza e richiede una analisi periodica ed accurata della lesioni melanocitiche del paziente. Abbiamo sviluppato una soluzione hardware e software chiamata Mole Mapper per assistere i deramtologi durante l'intero processo di diagnosi. L'obiettivo è permettere un incremento dell'accuratezza della diagnosi velocizzando al contempo l'intero processo. Tali caratteristiche si sono ottenute grazie ad un'analisi automatica delle immagini dermatoscopiche che individua ed evidenza al dermatologo le informazioni più significative. In questa tesi presentiamo 3 principali algoritmi che sono stati implementati in Mole Mapper:

Una robusta segmentazione di lesioni melanocitiche, che risulta il punto di partenza di ogni altro algoritmo di elaborazioni di immagini e permette l'estrazione di informazioni utili riguardanti la forma e la dimensione delle lesioni. Tale algoritmo supera in accuratezza e velocità lo stato dell'arte attuale, con una precisione paragonabile ad un dermatologo esperto ed un tempo di esecuzione compatibile con l'elaborazione video realtime;

Un algoritmo di depilazione digitale, che garantisce miglior precisione e robustezza agli altri algoritmi di elaborazione di immagini a fornisce al dermatologo un immagine priva di peli da impiegare nel processo di valutazione. La nostra proposta supera l'accuratezza dello stato dell'arte richiedendo solo una frazione del tempo di esecuzione, tanto da poter essere integrata su dispositivi mobili all'interno di una GUI interattiva.

Un algoritmo di registrazione, per studiare l'evoluzione delle lesioni nel tempo evidenziando ogni possibile anomalia o variazione. Data la mancanza di un approccio standard al problema, abbiamo caratteriizzato gli obbiettivi ed i vincoli a cui sottostare proponendo quindi un approccio con un tempo di esecuzione compatibile con le necessità del Mole Mapper ed un accuratezza paragonabile a quella di un operatore umano.

Chapter 1

Introduction

Malignancies of the skin are among the most common cancers known to man [1]: *Between 40 and 50 percent of Americans who live to age 65 will have a skin cancer at least once* [2]; *more people have had skin cancer than all other cancers combined* [3].

Melanoma accounts for less than 2% of all skin cancer cases, but causes the vast majority of skin cancer deaths [4]; moreover, its incidence is growing rapidly worldwide [5].

An early diagnosis is crucial to increase the survival rate since the excision of thin or in situ melanoma offers the possibility of mortality reduction [6]. At the same time, some aggressive forms of melanoma can lead to a very low survival rate within three months from their appearance [7] so very frequent check-ups may be necessary.

Periodic screening has proven [8] to be an effective way to decrease the unfavorable prognosis but, unfortunately, public infrastructures encounter difficulties in screening the population at risk at sufficiently high frequencies and the cost of private infrastructures is prohibitive for many subjects. The problem is that dermatologists and their time can be considered a limited resource that is not sufficient to satisfy the size of the population in need of it. An effective solution is to try to boost the dermatologists' performance, increasing the number of visits performed per day while at the same time maintaining a high standard of accuracy. Additionally, other types of screening which require minimal or no interaction by the dermatologists can be provided.

On this basis, the project CiS was born, with the purpose of providing a

modular and expandable system to empower the individual dermatologist's performance and allowing at the same time a long-term and continuous remote monitoring of the patient.

Section 1.1 gives an overview of the concept of the melanocytic lesion and the peculiarities of its most dangerous malignant form: melanoma. It also describes a common and effective in vivo examination technique (i.e. dermoscopy) and illustrates the typical diagnosis process.

Section 1.2 presents the structure of the CiS project which is composed of three main products: Mole Mapper, Full Body Scanner and Personal Screener.

Chapter 2 describes the key product of the CiS ecosystem: the Mole Mapper. It is designed to assist the dermatologist during the working day, empowering his overall performance. Mole Mapper uses many different computer vision algorithms that help the diagnosis process. Three of these are notably important since they are the key elements of the entire system and are used by the other image processing components:

- Melanocytic lesion segmentation (Chapter 3) accurately identifies the lesion borders in a dermatoscopic image

- Virtual Shave (Chapter 4) performs the hair detection and removal, improving the accuracy and robustness of the subsequent image processing modules and providing a hair-free image to the dermatologist

- Dermatoscopic images registration (Chapter 5) remaps two different images of the same lesion to the same coordinate system, allowing an effective analysis of the lesion's evolution.

1.1 Melanocytic Lesions

A melanocytic lesion is an anomaly in the skin formed by the proliferation of melanocytes. It is also often named melanocytic nevus and is equated with the term mole by some sources. Since the nomenclature can be slightly different in different sources, we define a melanocytic lesion as any benign or malignant anomaly caused by an atypical proliferation of melanocytes. We use the terms nevus (derived from the Greek word meaning nest) and mole as synonyms.

Nevi can be primarily classified depending on when they appear, which skin level is involved, and their malignancy.

Nevi can be congenital if present at birth or acquired if they appear during the patient's life. Nevi developed during the embryonic stage or in the first two years of life are both considered congenital since they are histologically equivalent. These lesions generally grow accordingly with the growth of the child and can sometimes reach a size greater than 20 cm in diameter. During the aging process they can often develop hair or become thicker. Acquired nevi often appear during childhood. The formation is generally related to solar radiation and genetic causes. The appearance of these lesions can be different depending on the melanocytes' depth and the specific cellular type.

Regarding the skin level location, if the nevus is confined to the dermoepidermal junction, the lesion is referred to as *junctional nevus* and it is generally flat and brown to black. It is an *intradermal nevus* if it is located in the dermis only; in this case, it is generally raised and often not pigmented. Finally, a nevus can be present in the epidermis and dermis, and is known as a *compound nevus*; it is generally slightly raised and brown to black.

Malignant skin lesions are the most common human cancer. In benign tumors, the proliferation is composed of well-differentiated cells and has limited growth. In particular, benign moles generally show very little change and remain almost static for years. Conversely, in malignant tumors the cells are undifferentiated and the growth is uncontrolled. Skin cancer which forms from melanocytes is called melanoma; this is the skin tumor that is responsible for most skin cancer deaths.

1.1.1 Melanoma

Melanoma (from the Greek melas, meaning "dark") is a malignant tumor which forms from melanocytes. It can develop from a pre-existing mole (ex nevo), or having an independent existence (de novo). According to the SEER Cancer Statistics Review [5], a new melanoma is diagnosed every 7 minutes in U.S. and a melanoma-related death occurs every 54 minutes. Worldwide, more than 130,000 new cases are reported each year [6]. The incidence is constantly growing (see Figure 1.1) and has doubled in the last 30 years [5]. Melanoma is the most aggressive form of skin cancer and has a high potential of metastatic spread.

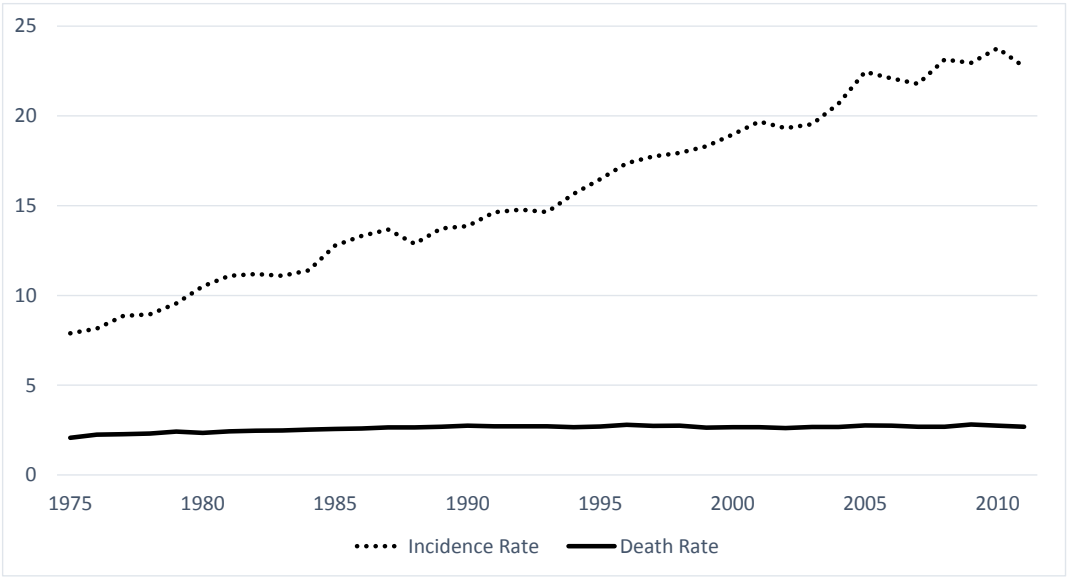


Figure 1.1: Melanoma annual death and incidence rate per 100,000 U.S. standard population.

Therapy options are limited in advanced disease, whereas an early diagnosis and prompt excision grant a good prognosis: the 5-year survival rate is above 95% for early diagnosis (localized stage) versus 15.7% for metastasized melanoma (distant stage) [9]. For that reason, a massive screening process (secondary prevention) can effectively increase the survival rate, as shown in the world's largest screening project (Skin Cancer Research to Provide Evidence for Effectiveness of Screening in Northern Germany) [8].

1.1.2 Dermoscopy

Dermoscopy, or epiluminescence microscopy, is a non-invasive, in vivo technique for the microscopic examination of melanocytic lesions. It effectively enhances melanoma detection and decreases the number of unnecessary excisions [10] [11] [12].

Dermoscopy is performed using an instrument called a dermatoscope, composed of a magnifier (typically providing a 10X zoom) and an incident light source. Generally, a liquid medium is used between the instrument and the skin, which allows the inspection of subcutaneous features of skin lesions reducing the skin surface reflections. More recent dermatoscopes make use of polarized light to

handle the reflection problem.

Experience and specific training are mandatory for dermoscopy since its practice by untrained or inexperienced dermatologists has been proven to be no better than clinical inspection without dermoscopy [11]. It is important for the operator to recognize different dermoscopic features. A significant example is the 7-point checklist (see Table 1.1) for dermoscopic scoring of atypical melanocytic lesions [13].

Dermoscopic criterion	Definition
Atypical pigment network	Black, brown, or gray network with irregular holes and thick lines
Blue-whitish veil	Irregular, structureless area of confluent blue pigmentation with an overlying white "ground-glass" film. The pigmentation cannot occupy the entire lesion and usually corresponds to a clinically elevated part of the lesion
Atypical vascular pattern	Linear-irregular or dotted vessels not clearly seen within regression structures
Irregular streaks	Brown to black, bulbous or finger-like projections irregularly distributed at the edge of a lesion. They may arise from network structures but more commonly do not.
Irregular dots/globules	Black, brown, round to oval, variously sized structures irregularly distributed within lesion
Irregular blotches	Black, brown, and/or gray structureless areas asymmetrically distributed within lesion
Regression structures	White scar-like depigmentation and/or blue pepper-like granules usually corresponding to a clinically flat part of the lesion

Table 1.1: 7-point checklist for dermoscopic differentiation between benign melanocytic lesions and melanoma [13].

While traditional dermatoscopy empowers dermatologists with a more detailed view of lesions, it fails at providing methods for image acquisition and comparison, which is the goal of digital dermoscopy. In this field a lot of effort has been spent trying to automatically extract useful information on digitally acquired dermoscopy images. Frequently, the automated system aims to perform a full lesion evaluation discriminating pictures containing a malignant lesion from the common mole. Equally important are the approaches that try to extract

all the meaningful information from the images for boosting the dermatologists' performance.

1.1.3 Diagnosis process

Skin Type	Phenotype	Response to sun exposure
I	Pale white; blond or red hair; blue eyes; freckles	Always burns, never tans
II	White; fair; blond or red hair; blue, green or hazel eyes	Usually burns, tans minimally
III	Cream white; fair with any hair or eye color; quite common	Sometimes mild burn, tans uniformly
IV	Moderate brown; typical Mediterranean skin tone	Rarely burns, always tans well
V	Dark brown; Middle Eastern skin types	Very rarely burns, tans very easily
VI	Deeply pigmented dark brown to black	Never burns, tans very easily

Table 1.2: Fitzpatrick skin classification scale [14].

During a visit session a dermatologist has to analyze all the lesions on the patient's skin, classifying them as probably malignant or non-malignant. A regular diagnostic process can be conceptually divided into three main steps: first there is the analysis of the patient's history and risk factors, followed by the identification of the suspicious lesions and finally a detailed analysis of such lesions for estimating their malignancy. Since the standard visitation process has not been defined, these steps can in practice be executed in a different order or combined, but the conceptual subdivision still remains.

For the first step, general information about the patient's medical history and his individual risk has to be evaluated. The patient's age, gender, personal and family history of skin cancer, and genetic and environmental risk factors have to be taken into account. The genetic factors include the classification of photosensitivity using the Fitzpatrick standard classification (see Table 1.2), the number and the aspect of congenital and acquired nevi and the personal history of any

previous melanoma or a positive case in first-degree relatives. Environmental risk factors are primarily related to UV radiation exposure. This factor is important not only as a risk factor for the development of the melanoma, but also for non-melanoma skin cancers, such as squamous cell carcinoma and/or basal cell carcinoma in whites. [15]

For the second step a detailed visual examination of the entire body including the nails, the hairy scalp , the soles and the visible parts of the oral and genital mucosa is required. Specific attention needs to be paid when the patient notices the development of new lesions or changes in pre-existing ones. At this stage the ABCDE [16] rule can serve as a clinical guideline to distinguish between benign and early malignant lesions during the examination with the naked eye:

A Asymmetrical Shape: Melanoma lesions are typically irregular, or not symmetrical, in shape. Benign moles are usually symmetrical.

B Border: Typically, non-cancerous moles have smooth, even borders. Melanoma lesions usually have irregular borders that are difficult to define.

C Color: The presence of more than one color (blue, black, brown, tan, etc.) or the uneven distribution of color can sometimes be a warning sign of melanoma. Benign moles are usually a single shade of brown or tan.

D Diameter: Melanoma lesions are often greater than 6 millimeters in diameter - yet many melanomas present themselves as smaller lesions, and all melanomas are malignant on day 1 of growth.

E Evolution: Any change – in size, shape, color, elevation, or another trait, or any new symptom such as bleeding, itching or crusting – points to danger.

The first 4 points (ABCD) can easily and quickly be evaluated by an experienced dermatologist, whereas the evolution aspect frequently relies only on the information provided by the patient. At the end of this step all the suspicious lesions are identified.

In the last step every suspicious lesion has to be carefully evaluated, generally using a dermoscopy analysis. In addition to the ABCDE rules, specific lesion features need to be analyzed. A standard example is the 7-point checklist (see Table 1.1) for the dermoscopic scoring of atypical melanocytic lesions [13]. The

analysis of a lesion includes the study of its appearance, the evaluation of its evolution and a comparison with the other lesions.

Our goal is to try to enhance the accuracy and speed of entire diagnosis process using a set of tools that boost the dermatologist's performance providing an effective user interface and many tools of (semi)-automated analysis.

1.2 Cutis in Silico

Cutis in Silico (CiS) is a collaborative project between the Department of Information Engineering, University of Padova, and the Dermatology Unit, School of Medicine. The project focuses on the design of a fast, accurate, and usable system for mole mapping and computer assisted melanoma screening.

CiS is designed as a modular, expandable system. Its modularity enables adoption by a wide range of users, from small consulting rooms to thousand-patient clinics. It is designed to empower the single dermatologist with tools for higher quality and productivity, as well as boosting efficient teamwork. It is a complete package to assist the doctor throughout the whole course of the visit, and more importantly allowing long-term, continuous remote monitoring of the patient.

The architecture of CiS comprises three main components which are autonomous devices designed individually for a particular phase of visit and its follow-up. The components are named Mole Mapper, which is a support tool for dermatological visits, Full Body Scanner, which is a complete figure photo booth and Mole Mapper, and Personal Screener, which is a low-cost, portable instrument allowing individual patients to acquire "at home" dermatoscopic images of lesions deemed suspicious by their dermatologist.

Even though the three main elements of CiS can be used independently, they can operate together with a high level of synergy. A self-synchronizing centralized data management system offers complete and up-to-date information from all devices, even across multiple laboratories and multiple dermatologists visiting the same patient.

The following is a brief description of the three CiS components:

Mole Mapper

The first of the three components of the CiS platform, the Mole Mapper is the core of the system. It is designed to be used by dermatologists during the typical working day and it is used as a main hub for the other two subsystems. The Mole Mapper prototyping phase has been completed and it is currently being used in a clinical test at the Dermatological Clinic of Padova. A detailed description of Mole Mapper is provided in the following chapter (Chapter 2)

Full Body Scanner

The Body Scanner can be seen as a subsidiary asset to the Mole Mapper, although it is intended as a stand-alone device. It takes high resolution photographs covering almost the entire area of the patient's body in a fraction of a second; from these, it then reconstructs an accurate 3D model of the body surface. Its first concept was designed in 2012 [17].

A considerable overhead in each visit is getting the patient naked and obtaining clear images of his body. Nevertheless, it is unavoidable in order to take reference images of the skin for two main reasons. First, it is essential to precisely document the position of pathological and suspect moles in order to avoid misunderstandings when communicating an excision to the surgeon. Second, one of the main clues that leads to melanoma diagnosis is the appearance of a new macula on the skin (melanoma de novo); it is estimated that the incidence of cutaneous melanoma developing from a pre-existing mole is as low as 20% of all cases (melanoma ex nevo). It is therefore essential to detect accurately and in advance the appearance of new lesions - which requires both a historical archive of reference images, and a precise comparison algorithm. [18] With the Body Scanner it is possible to automate image acquisition and the historical comparison of almost 100% of the patient's skin surface, while reducing the time costs of taking the full body images manually.

Personal Screener

The last component of the CiS platform is the Personal Screener. The Personal Screener is used by patients to keep track of the evolution of their moles in a follow-up program with minimal expense. Comprised of a software element and

a dermatoscopic lens, the Personal Screener kit contains an easy-to-use index of interested moles, and allows photograph management, as well as the basic automated evaluation of mole image parameters, possibly notifying the user if urgent dermatologic consultation is suspected necessary. All patients, dermatologists and clinical institutions will benefit from the use of the Personal Screener. Patients will be able to monitor the condition of their moles in between follow-up visits. Dermatologists will gain precious documentation of the evolution of lesions. Healthcare institutions will be able to match the optimal time resolution for screenings. Personal Screener provides end users with more frequent screening at a modest price, filling the gap between follow-up visits, and increasing the chances of early melanoma detection and favorable prognosis.

Chapter 2

Mole Mapper

Mole Mapper is the core of the CiS ecosystem. It is designed to be used primarily by dermatologists, supporting them during the typical working day. It also serves as a main interface and processing hub for the data produced by the other two subsystems: Full Body Scanner and Personal Screener.

A typical session of use of the Mole Mapper spans a single visit, from just before the patient enters the office, to immediately after the visit report is printed. Mole Mapper can access patient files through a phonebook-like archive, and can manage appointments via an inbuilt agenda, summarizing important information from the case history and recent updates on the patient's condition. It provides the means for taking full body images of the body surface, for guiding the dermatologist through the acquisition of dermatoscopic images of individual suspicious lesions on each body portion (a daunting task for patients sporting over a hundred lesions on their torso alone, for example). As soon as a full body image is acquired, whole body mole segmentation takes place; previously marked moles are then mapped to the new visual reference; the skin is finally scanned for the appearance of new moles. Dermatoscopic images of marked moles can also be compared to previous images of the same lesions, or to images of other lesions from the same patient. Clinical reporting is automated to reflect the status of the visit and the institutional standards.

The high level capabilities featured in the Mole Mapper reflect and expand on the skills of the dermatologist. The goal is not to replace the human operator,

but to help him to reach higher levels of accuracy, efficiency and confidence.

2.1 Hardware

Mole Mapper runs on Android devices. It is composed of a tablet and a detachable dermoscope.

Currently, two different prototypes have been manufactured and are undergoing clinical testing at the Dermatological Clinic of Padova. One is composed of an Asus Tranformer TF201 with a DermLite FOTO and the other one a Sony Xperia™Tablet Z with a DermLite DL3.

Asus Tranformer TF201 with DermLite FOTO

Figure 2.1: Prototype 1: Asus Tranformer TF201 with DermLite FOTO.

The first prototype was manufactured in March 2013 and is composed of an Asus Tranformer TF201 and a DermLite FOTO dermatoscope. The tablet provides good photograph quality and an additional docking station that adds 8 hours of battery life and keyboard functionality. The following are the technical specifications:

- Operating system: Android 4.0 (upgradable)
- Display: 10.1" LED Backlight WXGA 1280x800 display, Super IPS+, 10 finger multitouch support

- CPU: NVIDIA Tegra 3 Quad Core
- Memory: 1GB
- Storage: 32GB
- Camera: 8 MP Rear Camera with Flash, 1.2 MP Front Camera
- Battery: 18 hours pad with dock; 25Wh(pad) + 22Wh(dock) Li-polymer Battery
- Wireless Connectivity: WLAN 802.11 b/g/n@2.4GHz, Bluetooth V2.1+EDR
- Dimensions: 263 x 180.8 x 8.3 mm (19.4mm with dock)
- Weight: 586g (1123g with dock)

The DermLite FOTO integrates a cross-polarization system with 24 bright-white light emitting diodes (LEDs) and a four-element compound lens. The dermatoscope is permanently attached to a custom aluminum device provided with a quick-release adapter.

Sony Xperia™Tablet Z with DermLite DL3

Figure 2.2: Prototype 2: Sony Xperia™Tablet Z with a DermLite DL3.

This second prototype was manufactured in January 2014 and is composed of a Sony Xperia™Tablet Z and a DermLite DL3 dermatoscope. The photo quality and battery life are comparable with the previous solution. The CPU is faster,

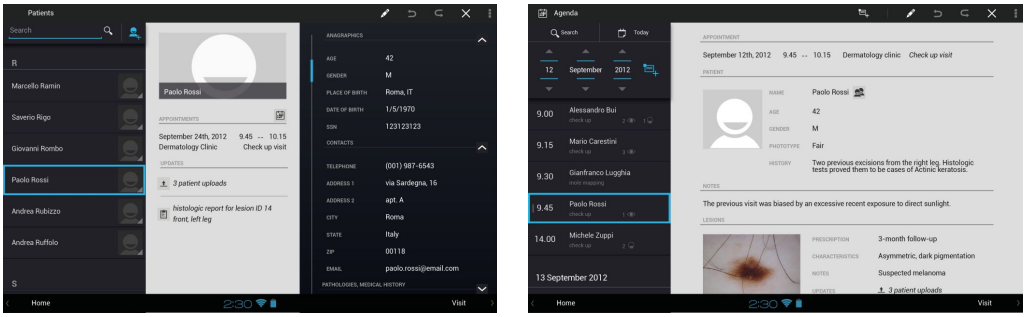
the memory is doubled and the device is thinner and lighter. The following are the technical specifications:

- Operating system: Android 4.2
- Display: 10.1" LED Backlight, 1920x1200, 10 finger multitouch support
- CPU: 1.5 GHz Qualcomm APQ8064+MDM9215M Quad Core
- Memory: 2GB
- Storage: 32GB
- Camera: 8.1 MP Rear Camera with Flash, 2.2 MP Front Camera
- Battery: 9 hours; 22Wh Li-polymer Battery
- Wireless Connectivity: WLAN 802.11 b/g/n@2.4GHz/5GHz, Bluetooth V4.0
- Dimensions: 266 x 172 x 6.9 mm
- Weight: 495g

The DermLite DL3 integrates a cross-polarization system with 21 bright-white light emitting diodes (LEDs) and a 25 mm four-element lens. It contains also 7 non-polarized LEDs for immersion fluid dermoscopy. The connection between the tablet and the dermatoscope was performed using a modified version of the DermLite Connection Kit for iPhone 5, allowing a consistent weight drop in comparison to the previous prototype.

2.2 Software

The software provides the support to the dermatologist for the entire visit session, starting from the agenda organization, continuing with the lesions acquisition and evaluation and ending with the report printing. In the next section is a list of the most significative section provided by Mole Mapper.



(a) (b)

Figure 2.3: Patients section screenshots. (a) Patient information summary. (b) Agenda.

2.2.1 Screens

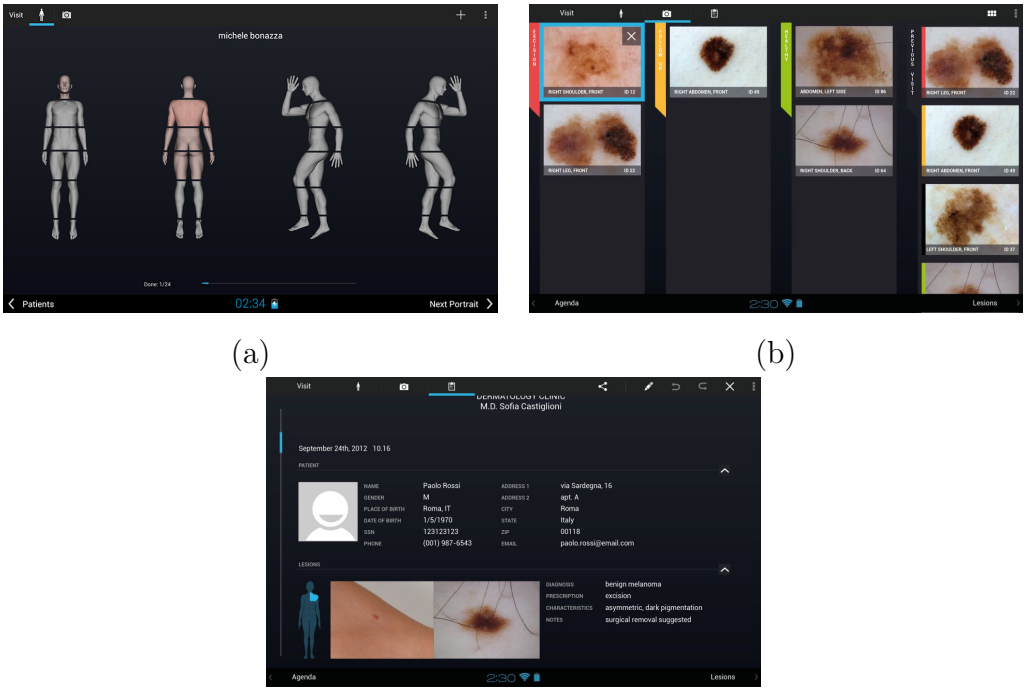
Patients

The Patients view group permits access to the patient's electronic health reports, and allows the user to gain complete information over the person's current and past medical condition. However, the Patients screen does not simply retrieve the information from the central database and report it plainly. Instead, it tries to highlight the most urgent and novel updates. Additionally, quick access to all contact information is provided. A report of all the past visits is shown, highlighting the most important and relevant information. Since all the data are stored in a centralized system, they can be easily updated remotely before the visit (e.g. by a secretary) to accelerate the data entry process.

Visit

The Visit section provides an overview to the dermatologist regarding the current visit status and allows for easy navigation between the different areas. It is divided into three main views: "portrait", "lesion" and "report".

The portrait outline (see Fig. 4.4a) is a view that gives information about the body coverage with full body photographs. The silhouettes, which match the gender of the patient, are divided into a standard set of portrait areas. Portraits acquired during this visit are fully colored, whereas partially colored portraits have been acquired in the past but not in the current visit. Finally, grayscale portraits are those that have never been acquired for this patient. Additional custom portraits can be defined in this view if some suspicious lesions are present



(a) (b)

(c)

Figure 2.4: Visit section screenshots. (a) Portrait outline. (b) Lesion overview. (c) Visit report.

in an area not covered by the standard portraits (e.g. behind an ear).

The lesion overview (see Fig. 4.4b) provides a synopsis of the suspicious lesions of the patient. It groups the lesions depending on the evaluation provided by the dermatologist and shows a queue of the current pending lesions. The pending group contains lesions that are in a follow-up status from the previous visit or those that have been marked as suspicious during the current visit but a final diagnosis has not been provided by the dermatologist. This view is useful not only as an overview of the current status but also for an easy evaluation of the lesions using the "ugly duckling" diagnosis rule: lesions markedly different from the remaining ones on the patient present a much greater risk.

The visit report view (see Fig. 4.4c) provides a preview of the final report, including photographs of the lesions and their locations. This can be used as a summary of all the visit aspects but also for easy navigation toward the most relevant lesion during the current visit.

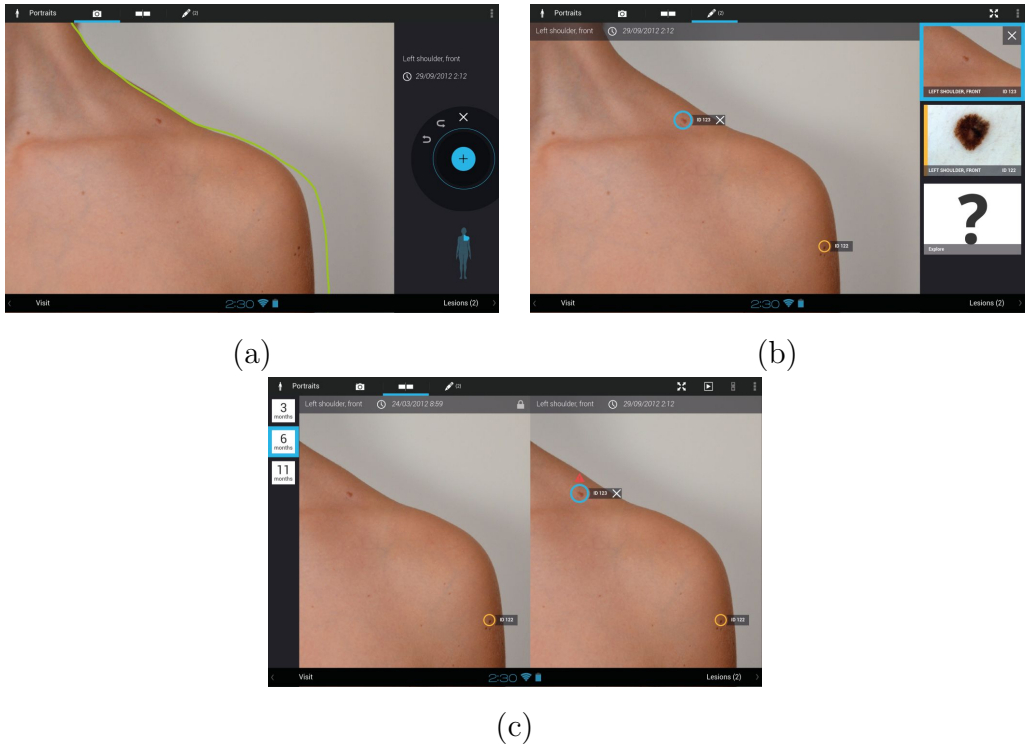


Figure 2.5: Portraits section screenshots. (a) Image acquisition. (b) Lesion marking. (c) Portraits comparison.

Portraits

The Portraits section provides a GUI for taking full-body photographs and performing their analysis. It allows the identification of new lesions and coarsely tracks the evolution of all of the lesions already present. It is composed of three screens: "camera" for taking new pictures, "mark" for dealing with information about the lesions and selecting the suspicious lesions and "comparison" for discovering the differences between different visits.

The "camera" screen (see Fig. 2.5a) allows the acquisition of a new picture for the current portrait. In order to facilitate a comparison, reproduction of the same pose in the new versions of the portrait is important. For this reason, the outline of the baseline version is displayed in overlay with the camera input as a guide, and is colored in such a way that it indicates when the tolerance between the poses is acceptable. The first time a portrait is acquired a standard outline is used, whereas for the following acquisitions the first picture taken is used for reference.

The "mark" screen (see Fig. 2.5b) shows the last acquired picture for the

current portrait and allows for interaction with the lesions in it. The dermatologist can "mark" the suspicious lesions that require further investigation simply by tapping on them. As a new nevus is marked, it is automatically segmented and given a unique identification number; a detail of the lesion in the context of the portrait is extracted and used as a placeholder image in the gallery, and will be later reused to locate the lesion. All of the lesions classified as malignant or in a follow-up status in the previous visit will be remapped to the new image, with a request to the dermatologist to provide feedback about them. In the Lesions section, the software will request that the dermatologist perform a dermoscopy analysis for all of the marked lesions.

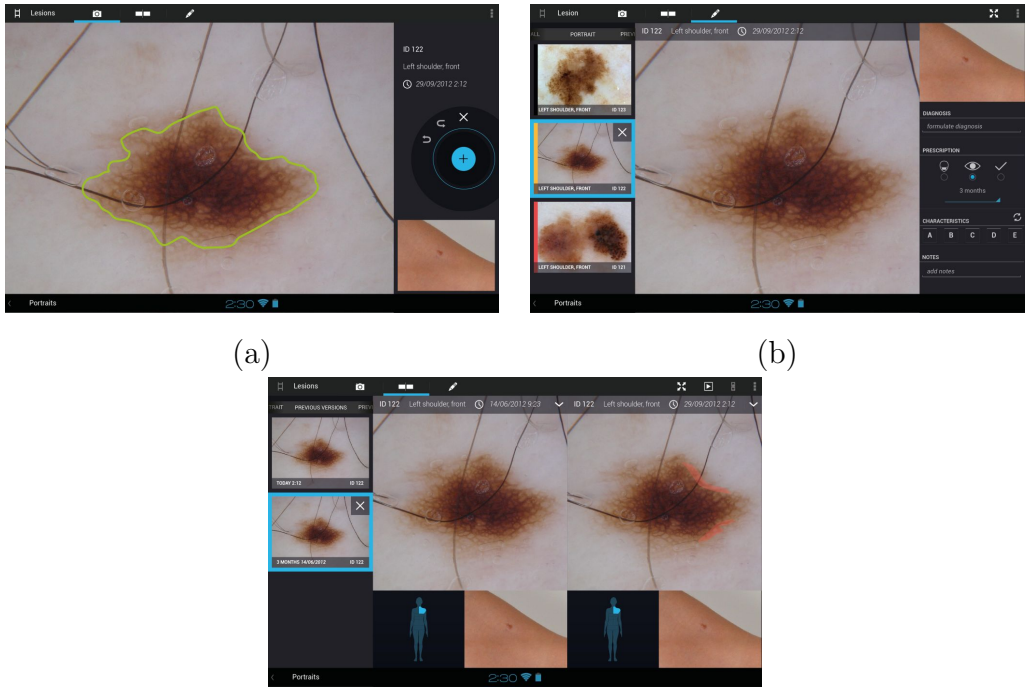
In the "comparison" (see Fig. 2.5c) screen the current portrait photograph is compared one-to-one with previous versions. This allows a much more detailed investigation about the lesion evolution and it is an invaluable aid in finding new lesions. Actually, the software performs the segmentation of all of the acquired images and performs a 1-to-1 remapping of all of the identified lesions. The lesions without a match are proposed to the dermatologist as possible new lesions or lesions with a significative change.

Lesions

The structure for the Lesion section is almost symmetrical to the Portraits section. It's used to perform a detailed evaluation of the lesions marked in the previous section. It is composed of three screens: "camera" for taking dermoscopic images, "characterization" for evaluating and classifying the lesions and "comparison" for analyzing the lesion evolution.

The "camera" screen (see Fig. 2.6a) allows the acquisition of a new dermoscopic image. Before the acquisition the previous image acquired is shown and during the acquisition process a small box depicting the lesion context in the portrait is available in the corner. This information helps the dermatologist to have a preliminary outlook on different lesion aspects and at the same time reduces the risk of mistaking two different close and similar lesions.

The "characterization" screen (see Fig. 2.6b) allows the analysis of all of the lesion information and the classification of each lesion. The lesion is automatically segmented and the attributes of many lesions are evaluated. This evaluation includes asymmetry, border peculiarities, color, size, the evaluation of the pig-



(a) (b)

(c)

Figure 2.6: Lesions section screenshots. (a) Image acquisition. (b) Lesion characterization. (c) Lesions comparison.

mented network, the vascular pattern, the presence of dots/globules and major changes from the previous visits. All of this information is saved in the lesion record and contextually shown to the dermatologist to help with the evaluation process. A prescription can be chosen with one tap as one of three radio buttons corresponding to excision, follow up or healthy nevus.

The "comparison" screen (see Fig. 2.6c) is close in spirit to the Portraits comparison, allowing a one-to-one similarity test between the dermoscopic images. Each pair of images is automatically registered and the differences are highlighted. Using this screen the dermatologist can perform a very detailed analysis on the lesion evolution between all of the available visits. Accessing this screen is not mandatory to perform considerations about a lesion evolution, since any major changes detected are automatically highlighted in the "characterization" screen.

2.3 Enhancing the visit process

The goal of the Mole Mapper prototype is to increase the dermatologist's accuracy during the visit, accelerating the diagnostic process at the same time. To find the best trade-off between the two aspects, a strict collaboration with the dermatologists' group has been established, alternating a development phase with the feedback from the practical verification.

foreach *portrait* **do**
| Take picture
end
foreach *portrait* **do**
| Mark suspicious lesions
end
Mount the dermoscope ;
foreach *marked lesion* **do**
| Take picture
end
foreach *marked lesion* **do**
| Classification
end

Figure 2.7: Standard visit workflow model on Mole Mapper.

Initially, a first round of interviews with five dermatologists was performed during a two-month period. Two of the dermatologists involved were amongst the most renowned professionals on a national scale. A first version of the prototype was developed and a clinical trial started. During this test, periodic interviews were scheduled for evaluating the weakest points and for discussing new features and required improvements. After two years of testing and development, we converged to a set of considerations and required features which are the cornerstones of our entire project. The most significant ones for this thesis are described below:

1-Lesion appearance evaluation The analysis of the lesion's appearance can be efficiently and effectively performed by an expert dermatologist, whereas the help provided by an automated system is pretty limited. Even if most of the effort in literature is focused on trying to emulate the dermatologist's behavior, having a tool that provides a full automated lesion appearance

evaluation does not increase the dermatologist's overall performance. However, automatic evaluation of the lesions' characteristics (e.g. asymmetry, border peculiarities, color, size, pigmented network peculiarities etc.) can be very useful for the following two reasons (2-3).

2-Ugly duckling criterion A comparison of the current lesion with the patient's other lesions is important due to a criterion called the "ugly duckling" rule: lesions markedly different from the remaining ones on the patient present a much higher risk. For this scenario an effective GUI together with a proper image analysis algorithm can provide a substantial boost to the dermatologist's overall performance.

3-Lesion evolution evaluation The evaluation of the lesion evolution can be a very demanding task when performed by a human operator. In fact, both short-term and long-term changes can represent interesting parameters and this requires the study of many different lesion pictures (when available). This is an area in which an automated evaluation system can provide a big boost to the dermatologist's performance.

4-New lesions identification The discovery of new lesions involves similar principles and considerations to those discussed in the previous point: recognizing the presence of a new mole on a patient with hundreds of lesions can be really challenging and time-consuming. A proper automated comparison can substantially increase the accuracy of this task, requiring only a negligible amount of time.

5-Defining a workflow The usage of a preconfigured workflow allows a more time-efficient approach, reducing the downtime and granting the execution of all of the necessary steps during the visit (see Fig. 2.7).

6-Portraits standardization As noted during user research, although the steps performed during a visit are homogeneous between most dermatologists, there is no publicly accepted standard in methodology for the acquisition of clinical data. In particular, non-standard body poses and non-standard subdivision of the body can make older images useless if the patient has been seen by multiple physicians with different personal conventions.

All of the previous elements can benefit from a congruous software architecture and an effective design of the interaction with the dermatologist. An analysis of these aspects together with an accurate dissertation about the Mole Mapper development process and the consequent design choices can be found in [19]. Additionally, the development of adequate image processing algorithms can lead to an exceptional improvement relative to elements 2-3-4. The three major image processing algorithms, whose development was crucial for the Mole Mapper automated analysis capability, are the main topic of this thesis and will be discussed in the following chapters.

Chapter 3

Melanocytic lesion segmentation

The first step in the visual analysis of a melanocytic lesion is segmentation [20], i.e. classification of all points in the image as part of the lesion or of the surrounding, non-lesional skin. While segmentation is typically studied in the context of automated image analysis, it is a first, necessary step even for human operators who plan to evaluate quantitative features of a lesion such as diameter or asymmetry – e.g. in epidemiological studies correlating those features to lesion malignancy [21].

The most important characteristic of a segmentation technique is accuracy, usually evaluated in terms of divergence from segmentations provided by one or more human "experts". The most widely used metric is simply the number of misclassified pixels normalized over the size of the lesion [22]. A crucial observation is that even expert dermatologists differ in their assessment of a lesion's border (see Fig. 3.1), since lesions are often fuzzy and there exists no standard operative definition of whether a portion of skin belongs to a lesion or not – dermatologists rely on subjective judgement developed over years of practice. The area of the disagreement region is typically $10 - 20\%$ of that of the lesion itself [23] [24]; this is obviously the minimum divergence that an automated system can be expected to have when evaluated against human experts.

Another important characteristic of any automated segmentation technique is computational efficiency. A slow segmentation can make any system based upon it unsuitable for real-time diagnosis; this is particularly true for hand-held, portable systems with limited computational resources.

This chapter presents MEDS [25] (Mimicking Expert Dermatologists' Seg-

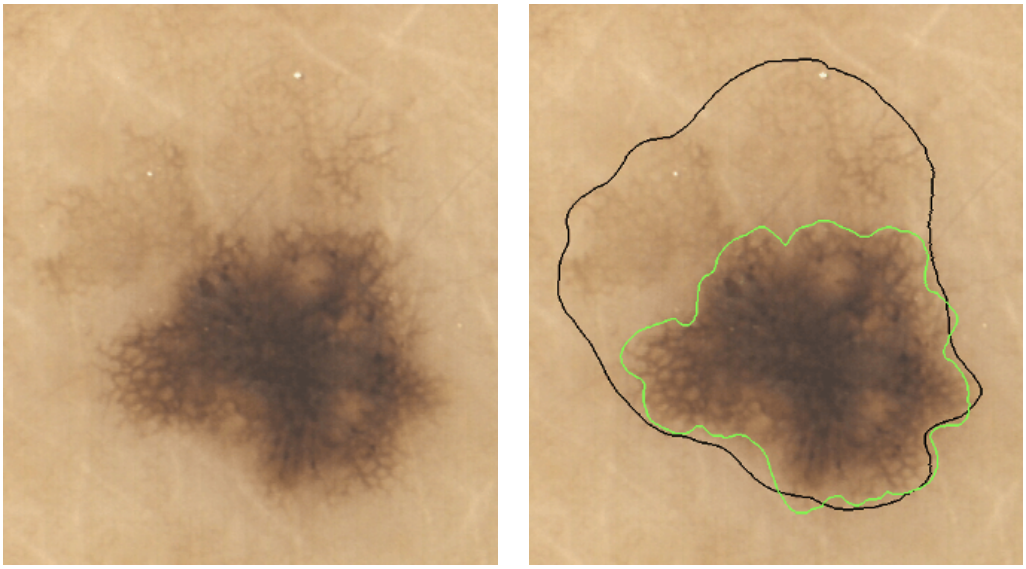


Figure 3.1: A dermatoscopically imaged melanocytic lesion (left) and two widely divergent segmentations obtained from two experienced dermatologists (right).

mentations), a novel technique for automated segmentation of dermatoscopically imaged melanocytic lesions. After Section 3.1 provides a brief review of the state of the art, Section 3.2 illustrates the details of MEDS. Section 3.3 then presents an experimental comparison of MEDS with other approaches, in terms of accuracy and computational efficiency, as well as an evaluation of its robustness to small image defects (such as air bubbles or unshaved skin), to illumination changes and to the inevitable deformations of the skin produced by a dermatoscope. Finally, Section 3.4 summarizes our results and discusses their significance.

3.1 Related work

We can separate into three main classes the numerous methods for lesion segmentation in dermatoscopic images (see [26] for an excellent survey).

The first class includes "minimal energy contours" techniques, that try to identify lesion *boundaries* through the minimization of a well-defined energy function. Commonly used energy functions consider edges and smoothness constraints, or statistical distributions over pixel intensities. A good representative of this class is Gradient Vector Flow (GVF) Snakes [27] [28]. The border identification accuracy of techniques in this class typically depends heavily on an initial segmentation estimate, on effective preprocessing (e.g. for hair removal) and on morphological

postprocessing [29] [30].

The second class includes "split and merge" techniques. These approaches proceed either by recursively splitting the whole image into pieces based on region statistics or, conversely, merging pixels and regions together in a hierarchical fashion. Representatives of this class include Modified JSEG [31], Stabilized Inverse Diffusion Equations (SIDE) [32], Statistical Region Merging (SRM) [33], Watershed [34]. Performance widely varies depending on a large number of parameters whose values must be carefully tuned [32] [23].

The third class of segmentation techniques for melanocytic lesions discriminates between lesional and non-lesional skin on the image's color histogram. After a preprocessing phase, these approaches classify each *color* as lesional or non-lesional. This separation is mapped back onto the original image, from which morphological postprocessing then eliminates small, spurious "patches". Simple thresholding techniques like Otsu's method [35] can provide accurate lesion segmentations in some cases, but in general lack robustness [36]; for example, they fail when lesions exhibit variegated coloring or low contrast with respect to the surrounding skin [37] [38]. More sophisticated approaches, such as Independent Histogram Pursuit (IHP) [39], Mean-shift [40] and Fuzzy c-means [41] [42] [23] [43] achieve greater robustness at the cost of increased computational loads. Our technique belongs to this third class.

3.2 Mimicking Expert Dermatologists' Segmentations in five stages

MEDS proceeds in five stages. The first (Subsection 3.2.1) is optional and simply preprocesses the image to rebalance its colors and/or to automatically remove any hair. The second stage (Subsection 3.2.2) reduces the dimensionality of the color space to 1 through Principal Component Analysis (PCA) of the color histogram. The third stage (Subsection 3.2.3) applies a blur filter to the resulting image to reduce noise. The fourth stage (Subsection 3.2.4) separates pixels into two clusters through a novel thresholding algorithm that is the heart of our technique and mimicks the cognitive process of dermatologists; this effectively partitions the original image into regions corresponding to lesional and non-lesional skin. The fifth stage (Subsection 3.2.5) morphologically postprocesses the image to remove

spurious "patches" and to identify lesional areas of clinical interest; it does so through a novel border detection scheme that appears at least 30% faster than the fastest existing schemes, and that may thus be of independent interest.

3.2.1 Preprocessing

Hair represents a common obstacle in dermatoscopic analysis of melanocytic lesions [44] [45]. Although our approach is relatively resilient to the presence of hair (see Section 3.3), in some cases automated hair removal significantly improves the final result. Thus, when necessary, we perform automated hair removal with VirtualShave [46].

We have also observed that, although our approach works well with any illumination that is reasonably balanced (more specifically, where a white object has Red, Green and Blue values all between 192 and 255), a cast with a strong Red component can significantly worsen the quality of the segmentation, whereas a cast with strong Blue, moderate Green and weak Red components can marginally improve it (again, see Section 3.3). This color balancing can be achieved either physically through the use of appropriate optics, or digitally by simply modifying "on the fly" the RGB values of each pixel the first time it is read from memory (an operation that takes negligible time).

3.2.2 PCA in Color Space

PCA [47] is a standard tool for statistical analysis of observations in a multi-dimensional space. We employ PCA to cluster the colors of the image into two classes according to their projection on the first principal component of the color histogram (where each point in the RGB space has a "mass" equal to the number of pixels with that color). Using only one dimension runs against the common wisdom of melanocytic lesion segmentation through PCA: virtually all previous work suggests one should use at least two.

In practice, we perform PCA on an m-pixel RGB image. We compute the 3×3 covariance matrix $\mathbf{C}$ as $\mathbf{M}^T\mathbf{M}$, where the ith row $\mathbf{m_i} = \langle r_i g_i b_i \rangle$ of the $m \times 3$ matrix $\mathbf{M}$ represents the three color components of the ith pixel, each component normalized by subtracting the mean value of that color in the image. Effectively

we have:

$$\mathbf{C} = \frac{1}{m} \sum_i \mathbf{m_i}^T \mathbf{m_i} \tag{3.1}$$

so that $\mathbf{C}$ can be easily computed by "streaming" the image pixel by pixel, subtracting the mean R, G, and B values, computing the 6 distinct products of the pixel's color components, and adding each of those products to the corresponding product for all other pixels (note that $\mathbf{C}$ is characterized by 6 elements rather than 9 since it is symmetric). Then, we compute the dominant eigenvector of $\mathbf{C}$, i.e. the first principal component of $\mathbf{M}$. This takes a negligible amount of time since it only requires computing the roots of a 3rd degree polynomial (the characteristic polynomial of $\mathbf{C}$) and inverting a 3×3 matrix. Finally, we project each row of $\mathbf{M}$ onto the principal component obtaining a one-channel grayscale image. Again, this can be achieved by "streaming" the image and performing only a few arithmetic operations for each pixel. Thus, the cost of the whole procedure is essentially that of scanning the image from memory three times (once for the average, once for the covariance, once for the projection).

We noticed dominant eigenvectors of different melanocytic lesion images were extremely close. In 60 images of different lesions from different patients, for any pair of dominant eigenvectors $\mathbf{v}$ and $\mathbf{u}$, we found $|\mathbf{v} \cdot \mathbf{u}| > 0.99$. We then decided to experiment with a simplified version of our technique, where instead of computing all eigenvectors of each image, one simply takes the (precomputed) average of the first eigenvector from a small training set of images. Throughout the rest of the article, we refer to this simplified version as *static MEDS*.

Section 3.3 shows that static MEDS still yields surprisingly good results while allowing significant speedup. Also, since the 1D color space on which the image is projected is independent of the image, static MEDS could simply employ (cheaper) grayscale image acquisition equipment paired with an appropriately tuned (physical) color filter – potentially allowing considerable cost savings when developing biomedical equipment to e.g. evaluate size, growth patterns or asymmetry of melanocytic lesions.

3.2.3 Noise Reduction

To reduce noise, we blur the grayscale image corresponding to the projection on the first principal component. More precisely, we apply a mean filter replacing the

value of each pixel with the average color in the 11×11 pixel square surrounding it. A naive implementation would require $11 \cdot 11 - 1$ additions plus one division for each pixel of the image. We reduce to 4 the number of additions required at each pixel by keeping track of the last computed values in a simple, auxiliary data structure (as in [48]). Furthermore, we perform each division by means of a multiplication followed by a shift; we have found this approach slightly more efficient than the one, based on a lookup table, employed by [48]. Our filter then requires only a single scan of the image and a handful of (non-floating point) operations per pixel, and is thus considerably faster than the fastest median filter implementations – while still providing comparable results in terms of final segmentation accuracy (see Table 3.1).

3.2.4 Color Clustering

Operating on the color histogram $h(\cdot)$ that associates to each color c the number of pixels $h(c)$ of that color, we separate colors (and thus pixels) into two clusters corresponding respectively to lesional and non-lesional skin. This stage, which is the heart of our technique and mimicks the cognitive process of human dermatologists, can be divided into three main phases. First, we apply to the histogram a square root operator, followed by a moving average operator over a window of 11 points. The square root operator enhances smaller values, which is useful when the percentages of lesional and non-lesional tissue differ widely. The averaging smooths out small fluctuations. More precisely, we have:

$$h'(x) = \sqrt{h(x)} \qquad h''(x) = \frac{1}{11} \sum_{y=x-5}^{x+5} h'(y) \tag{3.2}$$

Next, we find the positions M_ℓ, M_s of two local maxima in $h''(\cdot)$ that can be assumed as color "centres" of, respectively, lesional and non-lesional skin. Finally, we determine a threshold $F \in [M_\ell, M_s]$ separating the two clusters in the histogram.

The first centre M_1 corresponds to the global maximum in $h''(\cdot)$ (see Fig. 3.2). Note that M_1 cannot be classified as lesional or non-lesional until the second centre is found, since lesion area may be larger or smaller than non-lesional skin area. The second centre M_2 is computed as:

$$M_2 = \arg\max_x \left(h''(x)(h''(M_1) - h''(m_x)) \right), \; x \neq M_1 \tag{3.3}$$

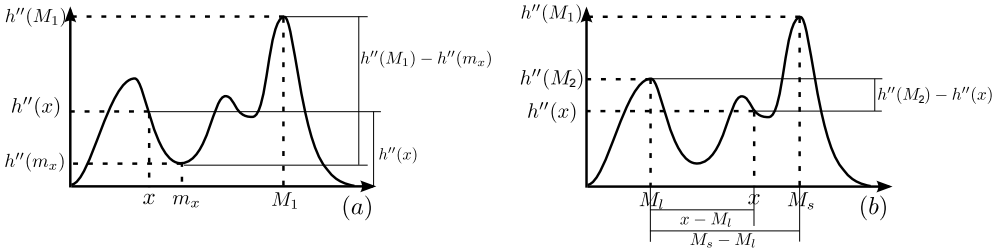


Figure 3.2: Partitioning of the color histogram into lesional/non-lesional colors.

where $h''(m_x)$ is the minimum of $h''(\cdot)$ between x and M_1. The two terms $h''(x)$ and $h''(M_1) - h''(m_x)$ in the maximized product favour, in the choice of M_2, a color that is "well-represented" (yielding a high $h''(x)$) and at the same time is "sharply separated" from M_1 (yielding a high $h''(M_1) - h''(m_x)$). This seems to accurately reflect the cognitive process of dermatologists.

To choose which of M_1 and M_2 should be classified as lesional, we simply assume lesional skin is darker, yielding:

$$M_\ell = \min(M_1, M_2),\ M_s = \max(M_1, M_2) \tag{3.4}$$

This assumption is satisfied by almost the totality of melanocytic lesions. Our technique could still be easily adapted to work in the extremely rare cases when this is not the case (such as amelanotic melanocytic lesions) by e.g. assuming that the lesion is entirely contained within the image and does not touch its borders – so that the color of the pixels on the image's borders is that of non-lesional skin [33].

Finally, we set the threshold between skin and lesion color:

$$F = \arg\max_x \Big(h''(M_2) - h''(x) \Big) \left(\frac{x - M_\ell}{M_s - M_\ell} \right)^\gamma \tag{3.5}$$

where $\gamma \in \mathbb{R}^+$ is the single "tuning" parameter of our technique – the smaller γ, the "tighter" the segmentations produced (see Fig. 3.3). Informally, the first term in the product favours as threshold a color that is *not* well-represented and thus yields a sharp separation between the two clusters. The second term, whose weight grows with γ, favours a color closer to that of non-lesional skin; this reproduces the behaviour of human dermatologists, who tend to classify as lesional regions of the image that are slightly darker than the majority of non-lesional skin, even when those regions are considerably lighter than the "core" of the lesion. Fig. 3.3 illustrates how the clustering results vary as γ increases from

0.8 to 1. On our dataset, we obtained good results for values of γ in $[1, 1.6]$ (see Section 3.3). Note that the fractional exponentiation in Equation 3.5 is carried out at most once for each of the 256 points of the color histogram, incurring an overall computational cost that is virtually negligible (again, see Section 3.3).

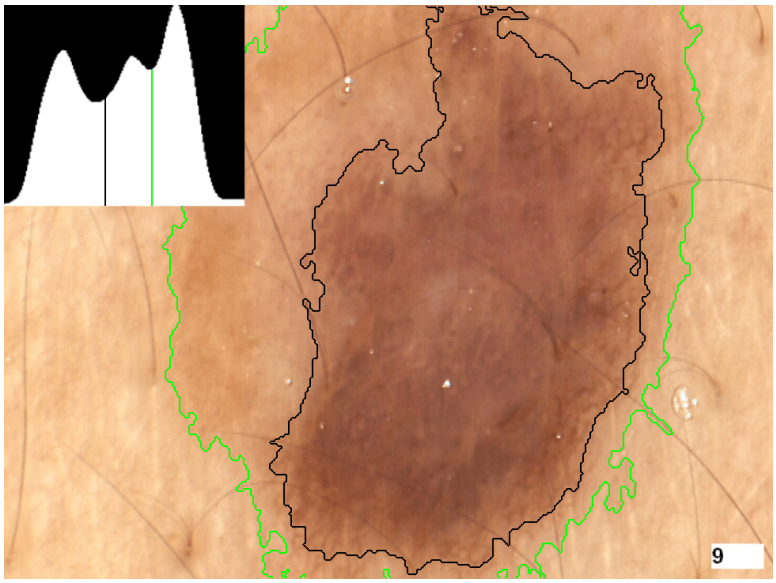
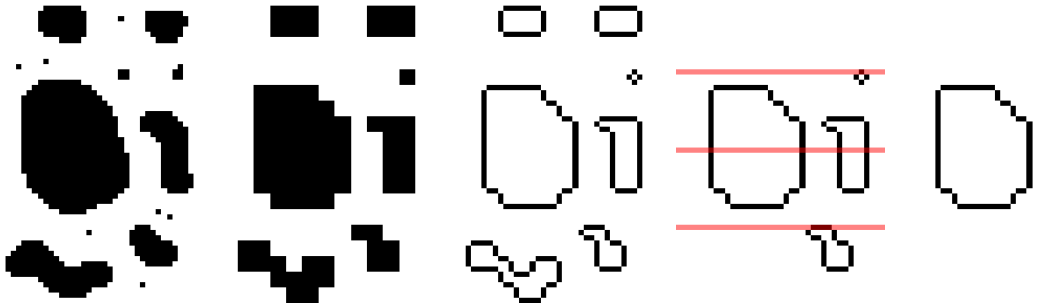


Figure 3.3: Identification of the separation point between lesional and non-lesional colors for $\gamma = 1$ (green) and $\gamma = 0.8$ (black). Lower values of γ yield "tighter" segmentations.

3.2.5 Postprocessing

Mapping the segmentation from color space back onto the original image produces a binary mask, where each pixel is classified as lesional or non-lesional. Two phases of postprocessing follow (see Fig. 3.4): first we "downsample" the image in order to easily identify the *boundaries* of each lesional component through a simplified (and faster!) version of the technique described in [49], then we remove all boundaries delimiting connected components that are "too small". This eliminates artefact "patches" due to individual pixels slightly darker or lighter than their neighbours, and identifies connected components classified as non-lesional but entirely surrounded by lesional pixels – such components usually correspond to air bubbles or lesion regressions and should be classified as lesional.

We now describe each phase in detail. Denote by $p_{i,j}$ the pixel located at row i and column j in an image, and by $v(p_{i,j})$ its value. For any pixel, we consider



(a) (b) (c) (d) (e)

Figure 3.4: The postprocessing stage. (a) Initial binary mask. (b) Binary mask after downsampling. (c) Boundary pixels. (d) d-rows. (e) Single boundary encircling "sufficient" area.

its 4- and 8-*neighbourhood* – informally, the 4 pixels adjacent to it horizontally or vertically, and the 8 pixels adjacent to it horizontally, vertically or diagonally. More formally, for each *internal* (i.e. non-edge and non-corner) pixel $p_{i,j}$ of an image:

Definition 1 *The* 4-neighbourhood *of* $p_{i,j}$ *consists of the 4 pixels* $p_{k,l}$ *such that* $|i-k|+|l-j|=1$.

Definition 2 *The* 8-neighbourhood *of* $p_{i,j}$ *consists of the 8 pixels* $p_{k,l} \neq p_{i,j}$ *such that* $|i-k| \leq 1$ *and* $|l-j| \leq 1$.

We deal with pixels on the edges or corners of the image by surrounding the image with a 1-pixel-wide strip of *non-lesional* pixels so that the pixels of the original image correspond to the internal pixels of the expanded image.

In the downsampling phase, we partition the (expanded) image into boxes of 3×3 pixels; each pixel in a box takes the value of the central pixel in the box:

$$v(p_{i,j}) \triangleq v(p_{k,l}) \text{ with } \quad k = 3\left\lfloor \frac{i}{3} \right\rfloor + 1 \ , \ \ l = 3\left\lfloor \frac{j}{3} \right\rfloor + 1 \tag{3.6}$$

Then, we identify the *boundary* pixels in the image:

Definition 3 *A* boundary *pixel is a lesional pixel whose* 4*-neighbourhood contains exactly* 3 *lesional pixels.*

The following theorem (whose proof can be found in the Appendix) establishes a crucial property of boundary pixels:

Theorem 1 *After downsampling, the* 8*-neighbourhood of any boundary pixel contains exactly* 2 *boundary pixels.*

By Theorem 1 then, if we view every boundary pixel as a vertex of degree 2 connected by an edge to its two adjacent boundary pixels, we obtain a set of disjoint cycle graphs, corresponding to the boundaries of all (putative lesional) connected components in the image. This makes it extremely easy to "walk" a boundary, starting from any of its pixels, following the edges between adjacent vertices in the corresponding graph.

The last phase of postprocessing computes the area of all connected components of "sufficient" height. A crucial notion for this phase is that of d-row:

Definition 4 *Consider an image of* r *rows, numbered from* 1 *to* r *starting from the top, and a parameter* d *(*$1 \leq d \leq r$*). We say the* i*th row is a* d*-row if* $i \bmod d = 0$.

Only boundary pixels belonging to a d-row serve as "starting points" to follow the corresponding boundary. Every component with height at least d then gets "caught", while smaller components *may* be missed (if no d-row intersects them – see Fig. 3.4(d)); but these "small" components are of no interest to us. d-rows allow considerable speedup as long as d is larger than $5-10$; while d values equal to (or smaller than) 5% of the image's height catch all lesions of clinical interest. Thus, we set d as 5% of the image's height.

From the boundary of a connected component, we easily compute the area: denoting by b_i the ith boundary pixel of the component on a generic row, the pixels of the component in that row are those between any two consecutive boundary pixels b_i and b_{i+1} with *odd* i. We then remove all boundaries delimiting areas smaller than one fifth that of the largest connected component: this takes care both of small dark patches in non-lesional skin, and of small light patches within a lesion.

Note that there are many known techniques to identify connected components in a binary image (e.g. [50] [51] [52]), but they are more computationally expensive than ours, requiring at least two scans of the image and/or additional data structures. In contrast, our technique makes a single sequential pass plus a small number of additional accesses to a limited number of pixels. Even the optimized,

single-pass approach of [52] requires approximately 30% more time than ours, plus additional effort to "match" portions of the lesion or of the skin that do not belong to the same connected component.

3.3 Experimental evaluation

We evaluated MEDS in terms of accuracy, computational efficiency and robustness. Subsection 3.3.1 briefly describes our experimental setup. Subsections 3.3.2 and 3.3.3 evaluate MEDS in terms of accuracy and computational efficiency, respectively, by comparing it to three different state-of-the-art techniques. Finally, Subsection 3.3.4 evaluates its robustness to illumination changes and to the inevitable deformations of the skin produced by a dermatoscope.

3.3.1 Experimental Setup

60 images of melanocytic lesions at 768 $\times$ 576 resolution were acquired with a Fotofinder digital dermatoscope [53]. 12 copies of each image were printed on 13cm $\times$ 18cm photographic paper. A copy of each image and a special marker pen were given to each of 4 "junior", 4 "senior" and 4 "expert" dermatologists (having respectively less than 1 year of experience, more than 1 year but no specific dermatoscopic training, more than 1 year and specific dermatoscopic training). Each dermatologist was asked to independently draw with the marker the border of each lesion. The results were scanned and realigned to the same frame of reference, and the contours provided by the markers were then extracted and compared – identifying, for each pixel of each original image, the set of dermatologists classifying it as part of the lesion or of the surrounding non-lesional skin. This "pen-and-paper" approach aimed at maximizing the comfort of dermatologists, thus minimizing the noise in border localization caused by the use of unfamiliar software drawing tools [24].

We implemented our technique in Java and tested it on three different platforms: a Samsung Galaxy S smartphone with a 1 GHz ARM Cortex A8 processor, an ASUS Transformer Prime tablet with a 1.3 GHz Nvidia Tegra 3 processor, and a desktop PC with a 3.07 GHz Intel Core i7-950 processor. To provide a clearer evaluation of the strengths and limitations of our technique, none of our tests made use of the optional digital hair removal phase (see Subsection 3.2.1).

We compared MEDS with three different state-of-the-art approaches, selecting a representative technique for each of the three classes introduced in Section 3.1 and privileging those with publicly available implementations. Since simpler minimal energy contours methods (like GVF Snakes) tend to sport poor accuracy [29] [30], for the first class we tested EdgeFlow [54], that includes a texture component to yield more robust edge detection. For the second class, we tested Statistical Region Merging (SRM) [33]. For the third class, we tested a Java implementation of 2D-PCA [23].

Since SRM and EdgeFlow are written in C (usually more efficient, but less portable than Java) we could test them only on the Core i7 platform. Also, SRM does not work properly on lesions adjacent to the image's borders, so we did not test it on any such images. This yielded a reduced dataset of 40 images, on which our own technique's segmentations were more accurate than on the full dataset (see Subsection 3.3.2) – so we effectively gave SRM an advantage by running it on an "easier" dataset. EdgeFlow produces a set of segmented regions, but does not include a decisional step to determine which regions should be marked as part of the lesion. Again, we made the comparison as biased as possible *against* our own technique, by assuming Edgeflow augmented with an "ideal" decisional step taking zero time and returning the set of regions maximizing segmentation accuracy (see Subsection 3.3.2).

3.3.2 Accuracy

We measured the accuracy of a generic segmentation S by comparing it to a "ground truth" reference segmentation R, and counting the number TP of true positive pixels (classified as lesion by both segmentations), the number FP of false positive pixels (classified as lesion by S but not by R), the number FN of false negative pixels (classified as lesion by R but not by S) and the number TN of true negative pixels (classified as lesion by neither segmentation). We then computed the *divergence of S from R* as:

$$d_s = \frac{FP + FN}{TP + FN} \tag{3.7}$$

i.e. as the ratio between the area of the misclassified region (FP+FN) and the area of the lesion itself according to the ground truth reference segmentation (TP+FN) [55].

We evaluated the different techniques by comparing their segmentations with those produced by the 4 expert dermatologists (see Table 3.1). MEDS obtained, on average, 12.35% disagreement with expert dermatologists. A slightly modified version of our technique, that we call *MEDS boost*, reduced the disagreement to 11.27% by enhancing the Blue and Green channels of the image. MEDS boost first rebalances the image's colors and then normalizes the mean value of each channel: in the preprocessing phase, the Red, Green and Blue values of each pixel are multiplied respectively by 0.02, 0.2 and 0.98, and each of the three color values is divided by the mean value of that color in the image before computing the covariance matrix. In this way, MEDS boost *adaptively* scales the variance of each channel, ensuring robust PCA and thus accurate segmentations even when the ratio between lesional and non-lesional pixels in the image is very low.

As mentioned in Section 3.2.3, performing noise reduction through a median filter (the entry "MEDS median filter" in Table 3.1) appears to produce virtually no improvement in accuracy compared to the much faster mean filter used by MEDS. Also, substituting our thresholding scheme with the classic method of Otsu [35] while leaving all the other stages of MEDS unchanged (the entry "MEDS Otsu's thresholding") produces a fair drop in accuracy, confirming the effectiveness of our specialized thresholding scheme.

In the spirit of [23], we also evaluated the 4 senior and 4 junior dermatologists using as ground truth the segmentations produced by the 4 expert dermatologists, and each expert dermatologist using as ground truth the segmentations produced by the remaining 3 expert dermatologists. The average divergence of junior dermatologists from the experts, of the senior dermatologists from the experts, and of the experts from the other experts, was respectively 17.24%, 13.57% and 10.40%. Thus, MEDS achieved a disagreement with expert dermatologists that was lower than that achieved by junior and senior dermatologists, and very close to the disagreement of expert dermatologists between themselves (see Fig. 3.5).

Results in Table 3.1 were obtained setting γ equal to 1, the value minimizing average disagreement with expert dermatologists on the entire dataset. To rule out the possibility of an excessively optimistic evaluation due to overfitting, we carried out 30 trials of random subsampling validation. In each trial, we randomly partitioned the 60 lesion dataset into a 30 lesion training set and a 30 lesion validation set, measuring average disagreement with expert dermatologists on

Table 3.1: Divergence d_s (average and standard deviation) from expert dermatologists in the segmentation performed by different dermatologists and automated techniques.

Group	d_s (avg)	d_s (std)
Experts	10.40%	6.86%
Seniors	13.57%	9.54%
Juniors	17.24%	15.53%
MEDS boost	11.27%	6.33%
MEDS	12.35%	6.98%
static MEDS	12.45%	7.16%
static MEDS w/o NR	13.44%	8.21%
MEDS median filter	12.26%	7.12%
MEDS Otsu's thresh.	14.52%	7.60%
2D-PCA	15.58%	7.19%
SRM	15.15%	8.65%
EdgeFlow	16.75%	8.06%

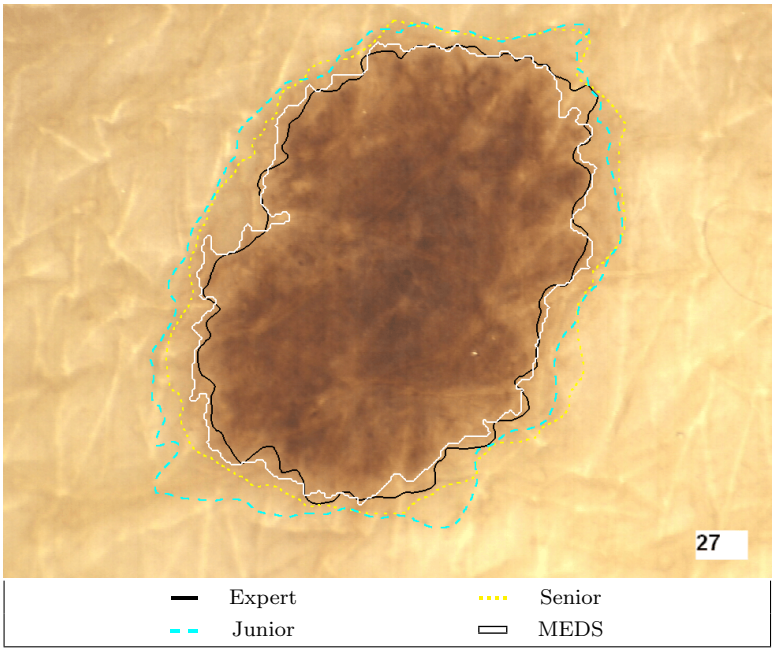


Figure 3.5: Melanocytic lesion segmentation performed by human dermatologists and MEDS.

the validation set using the value of γ that minimizes average disagreement on the training set. Fig. 3.7 shows the values of γ and of average disagreement for

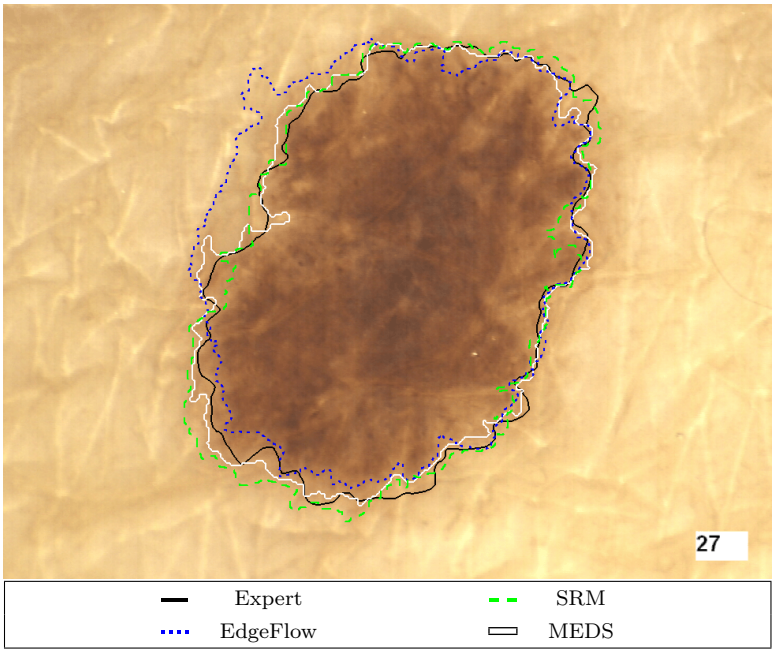


Figure 3.6: Melanocytic lesion segmentation performed by expert dermatologists, Statistical Region Merging (SRM), EdgeFlow and MEDS.

each trial; trials are sorted by increasing disagreement. In 26 out of 30 trials γ was in the interval $[1, 1.6]$; in the remaining 4 it was in the interval $[1.9, 2.6]$, yielding slightly higher disagreement. Average disagreement per trial ranged from 10.46% to 14.18% – for an overall average of 12.71%. One should note that our lesion dataset is highly inhomogeneous in terms of size, color, illumination, and presence of artefacts (e.g. air bubbles or hair).

Some simplifications of MEDS appear to incur only modest accuracy reductions. Static MEDS (Subsection 3.2.2) incurs a negligible 0.1% loss in accuracy; eliminating the noise reduction step (see Subsection 3.2.3) incurs a slightly larger 1% loss (see Table 3.1). Subsection 3.3.3 shows how these small accuracy losses can be traded for significant speedup. Again, these results do not appear biased by overfitting; in fact, a remarkably small training set seems sufficient to obtain a "good" estimate of the principal eigenvector. In 30 trials each involving a training set of only 10 images and a validation set of 50, the dot product between the average of the principal eigenvectors of the training set and the average of the principal eigenvectors of the validation set was never less than 0.99, making the modulus of the difference vector always less than 0.1. In those 30 trials the average divergence from expert dermatologists of static MEDS on the validation

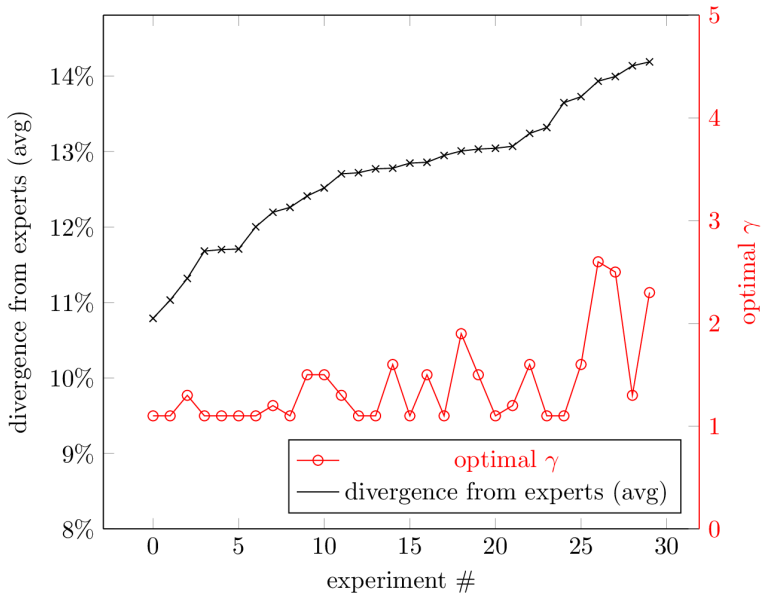


Figure 3.7: Disagreement of MEDS with expert dermatologists averaged over 30 random images, using the optimal value of γ obtained for the remaining 30 images, for each of 30 trials sorted by increasing divergence.

set (using the principal eigenvector computed on the training set) ranged between 11.24% and 13.19%, with an average of 12.34%.

All other automated techniques exhibited worse accuracy. EdgeFlow provided the worst results, with an accuracy comparable to that of junior dermatologists, despite our "generous" evaluation which, for each image, considered lesional the set of regions minimizing divergence from the ground truth (see Subsection 3.3.1). The accuracy of SRM, too, was worse than that of senior dermatologists, again despite a "generous" evaluation on the easier, reduced dataset (by means of comparison, MEDS improved its divergence from 12.35% to 11.87% when moving from the full dataset to the reduced one). Even 2D-PCA was less accurate than MEDS; this difference may be due in part to the fact that the second principal component introduces more noise than information, but is probably mostly due to our more sophisticated thresholding scheme.

3.3.3 Computational Resources

Our segmentation technique is extremely fast. Segmenting any one of our test images in memory took less than 0.02 seconds on the Core i7 desktop and only

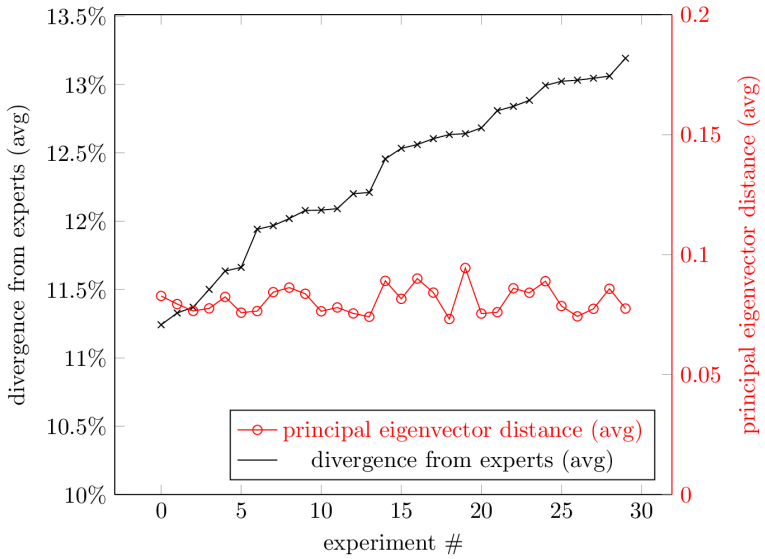


Figure 3.8: Disagreement of static MEDS with expert dermatologists averaged over 50 random images, using the average of the principal eigenvectors of the remaining 10 images, for each of 30 trials sorted by increasing divergence; and the corresponding distance between the average eigenvector of the 10 image training set and of the 50 image validation set.

0.7 seconds on the Galaxy S smartphone (this does not account for the possible cost of preprocessing with a hair-removing tool, or that of software color balancing – the latter being negligible anyway, as noted in Section 3.2). Table

Table 3.2: Execution time in milliseconds of MEDS, static MEDS with and without noise reduction, 2D-PCA, SRM and EdgeFlow on a desktop PC with an Intel Core i7-950 processor, on a Samsung Galaxy S phone and on an ASUS Transformer Prime tablet.

	Desktop	Smartphone	Tablet
MEDS	17	733	411
static MEDS	12	407	286
static MEDS w/o NR	7	185	120
2D-PCA	199	5986	2778
SRM	189	N/A	N/A
EdgeFlow	104789	N/A	N/A

3.2 shows how skipping some computation-intensive operations with marginal ef-

fects on accuracy can significantly lower execution time: static MEDS required $30-45\%$ less execution time than MEDS while providing virtually identical accuracy. Similarly, skipping the noise reduction phase (and thus worsening accuracy by a modest 1%) reduced execution time by $30-40\%$ (and by $50-55\%$ in the case of static MEDS). Fig. 3.9 shows the contributions of each phase to the total execution time.

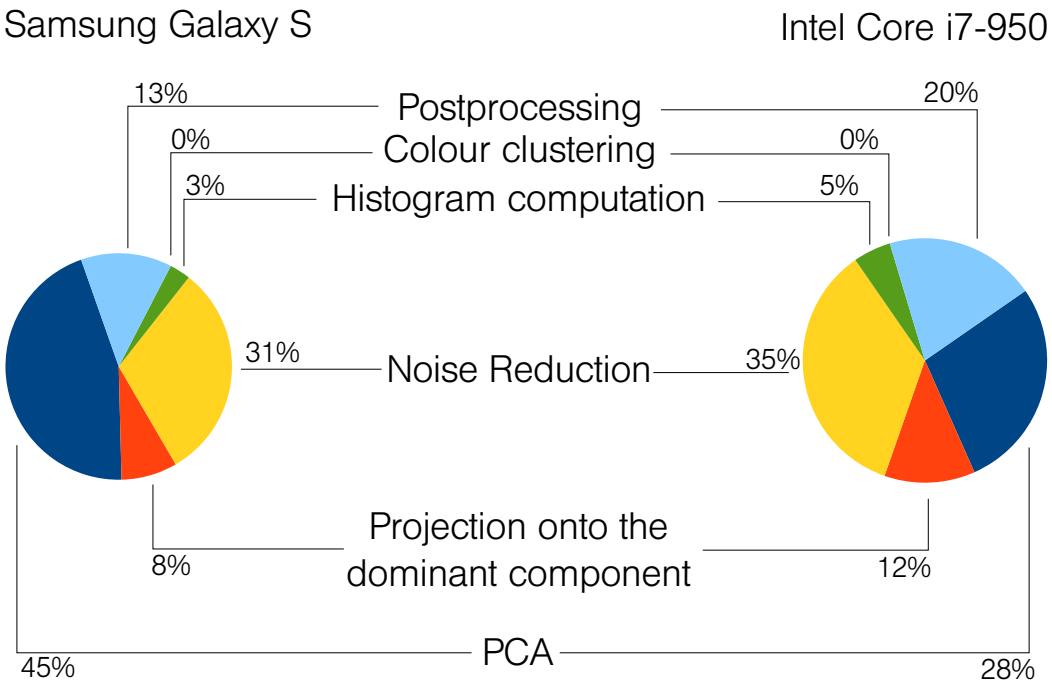
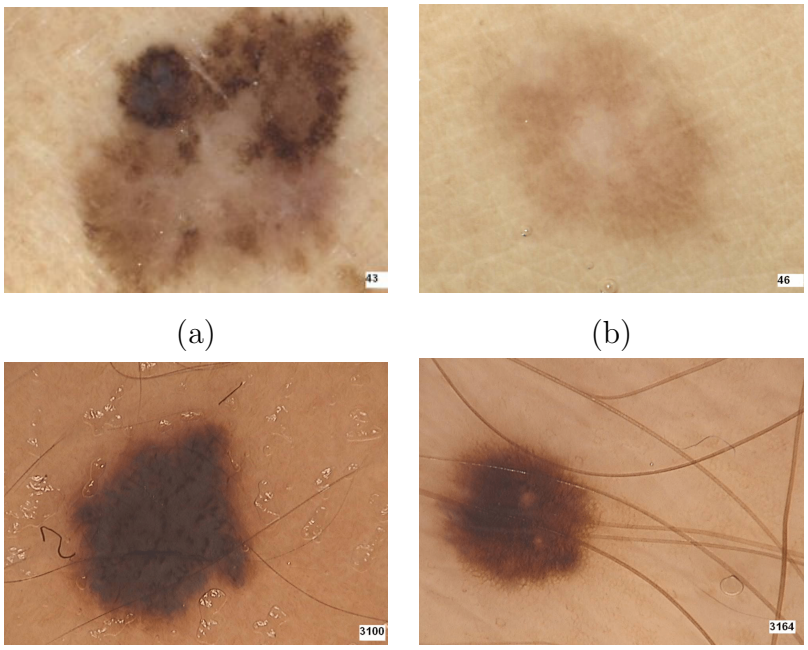


Figure 3.9: Time cost breakdown of MEDS on a Samsung Galaxy S cell phone and on a desktop PC equipped with an Intel Core i7-950 processor.

MEDS outperformed both SRM and 2D-PCA by over an order of magnitude in terms of running time; and EdgeFlow by several orders of magnitude (even though we "charged" EdgeFlow no time costs for the choice of the lesional region set – see Subsection 3.3.1). The main reason for the extreme computational performance of MEDS is that its 1D color histogram is processed very quickly: only a handful of operations are required for each of its 256 points, without any need of costly iterations. And since PCA, color histogram creation, and morphological postprocessing all boil down to "streaming" the image while performing a few simple operations on each of its pixels, the total cost of segmenting the image is essentially that of scanning it a few times.



(a) (b)

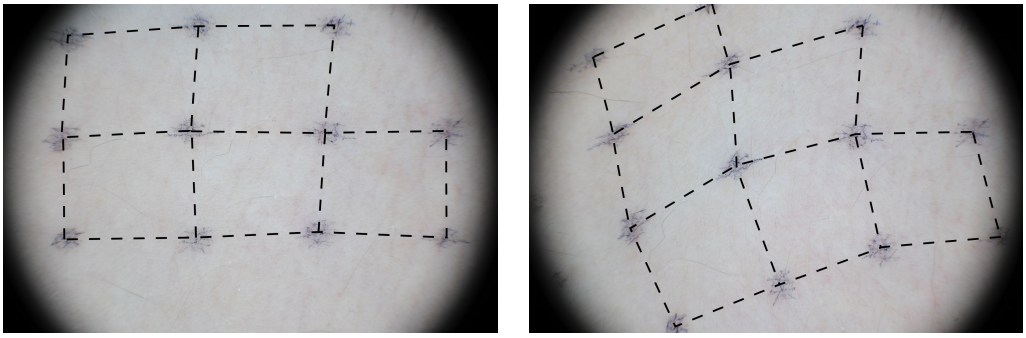
(c) (d)

Figure 3.10: Melanocytic lesions exhibiting inhomogeneous pigmentation (a), low color contrast against surrounding skin (b), air bubbles (c), hair (d).

3.3.4 Robustness

Some dermatoscopically imaged lesions are considerably harder to segment than others [29] due to intrinsic properties of the lesion (e.g. inhomogeneous pigmentation or low color contrast with the surrounding skin – see Fig. 3.10(a)-(b)) or to suboptimal image acquisition (e.g. presence of unshaved hair, air bubbles trapped in the anti-reflective gel, or shadows cast by the dermatoscope – see Fig. 3.10(c)-(d)). And different images of the same lesion, even taken within few seconds of each other and with the same equipment, zoom and framing, can present to the viewer considerably "different" lesions: it is difficult to guarantee consistent illumination, while even mild pressure from the dermatoscope can cause significant deformation of the skin. We assessed accuracy variations of our technique on different types of "difficult" lesion images, and segmentation reproducibility in the presence of illumination variations and of skin deformations.

As for accuracy variations, we identified 4 (non-disjoint) subsets of our dataset containing respectively 33 lesions with inhomogeneous pigmentation, 19 lesions with low color contrast against the surrounding skin, 35 lesions imaged with air bubbles, and 24 lesions imaged with unshaved hair. MEDS obtained an average



(a) (b)

Figure 3.11: Skin deformation due to the dermatoscope pressure. (a) and (b) depicts the same skin area acquired in two different dermatoscopic photos.

divergence from expert dermatologists of 12.55% for the inhomogeneous pigmentation subset, of 14.25% for the low contrast subset, of 12.19% for the air bubble subset, and of 11.14% for the hair subset.

To assess robustness in the presence of skin deformations, we deformed each image of our dataset with a combination of a rototranslation, a perspective distortion and a barrel distortion, trying to include all possible factors affecting an actual dermatoscopic image (and producing much more dramatic deformations than those observed in practice). We then measured the disagreement of the deformed segmentations produced on the *original* images by MEDS and by expert dermatologists, with the segmentations produced by MEDS *directly* on the deformed images. Average disagreement between the segmentations produced by MEDS on the original and deformed images was 3.17% (about 1% attributable to rounding in the deformation). Average disagreement with expert dermatologists was 13.07%, slightly higher than in the absence of deformations but still lower than the disagreement between expert and senior dermatologists.

As for sensitivity to moderate illumination variations, we considered 27 versions of each image obtained by independently reducing the R, G and B color values by 0%, 12.5% or 25%. The average and maximum divergence of MEDS between versions were 1.00% and 1.78%, and those of static MEDS 1.07% and 2.05%. The average disagreement from expert dermatologists was 12.50% for MEDS and 12.35% for static MEDS – virtually the same obtained with standard illumination. More significant illumination variations did produce more significant effects, in some cases (mostly involving a moderate to strong relative boost

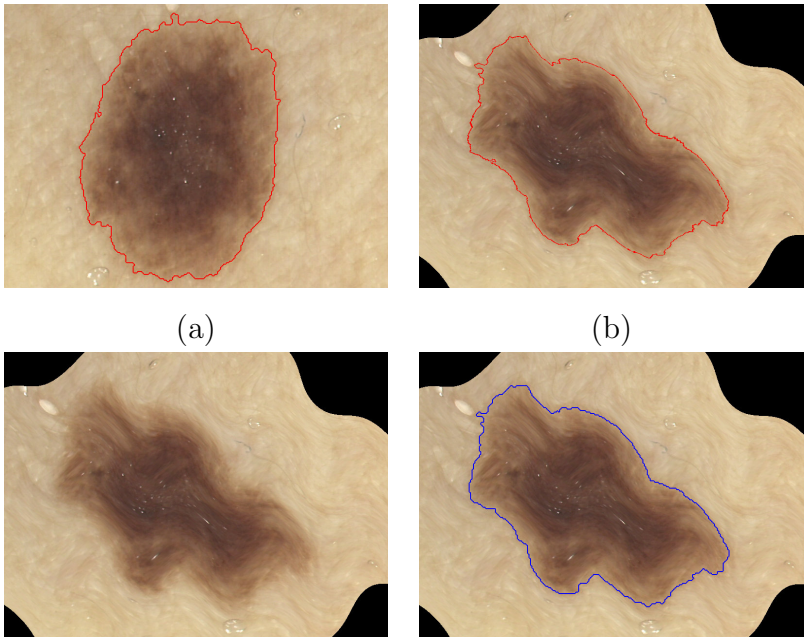


Figure 3.12: Deformation robustness test. (a) *Original* image. (b) Deformed segmentation produced on the *original* image. (c) *Deformed* image. (d) Segmentation produced on the *deformed* image.

of Blue, possibly combined with a weaker boost of Green) improving agreement with experienced dermatologists (see Section 3.3.2).

3.4 Conclusions

Our simple technique for melanocytic lesion segmentation is very accurate, thanks to its novel thresholding scheme which mimicks the cognitive process of human dermatologists. It appears more accurate than state-of-the-art techniques. Per-

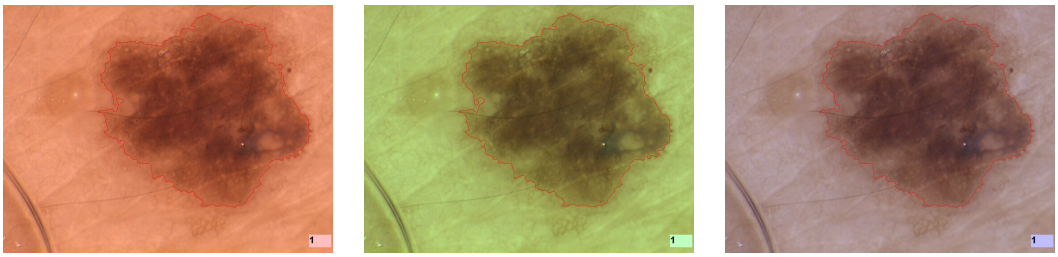


Figure 3.13: Segmentation produced on images with artificial illumination changes.

haps more importantly, it appears almost as accurate as *any* segmentation technique can be, since expert dermatologists disagree with it only slightly more than they disagree between themselves – and less than they disagree with dermatologists of little, or even moderate, experience.

At the same time our technique is *extremely* fast, in part due to a number of optimizations (such as our border identification scheme) that may be of independent interest. A Java implementation can segment a medium sized dermatoscopic image in the time required to simply scan the image a handful of times – a fraction of a second even on hand-held devices with modest computational resources. This is an improvement of an order of magnitude or more over the state of the art.

Finally, our technique is robust. It does not require careful hand-tuning; it sports a single parameter that controls how "tight" the segmentation is. It tolerates very well small photographic defects, such as small air bubbles or uneven lighting, and yields highly reproducible segmentations even in the presence of skin deformations or illumination changes. In fact, our technique is so robust that one can achieve almost as accurate results with a crude simplification which, instead of projecting the color space of each image onto its principal component, projects it onto a precomputed space independent of the image – allowing even faster processing, as well as use of (cheaper) monochromatic image acquisition equipment.

Appendix: Proof of Theorem 1

Theorem 1 *After downsampling, the* 8*-neighbourhood of any boundary pixel contains exactly* 2 *boundary pixels.*

Consider an arbitrary boundary pixel $p_{i,j}$, and denote by B the 3×3 box containing $p_{i,j}$. The downsampling phase ensures that pixels in B are homogeneous (either all lesional or all non-lesional); since $p_{i,j}$ is lesional by definition, B contains 9 lesional pixels. The intersection between B and $p_{i,j}$'s 4-neighbourhood contains either 2 pixels (i.e. $p_{i,j}$ is a vertex of B) or 3 pixels (i.e. $p_{i,j}$ is neither a vertex nor the central pixel of B). Note that $p_{i,j}$ cannot be B's central pixel, since its 4-neighbourhood contains 3 lesional pixels. We prove the theorem by analysing the two cases separately.

Consider the first case. For simplicity, we assume that $p_{i,j}$ is B's upper-left vertex (for the other vertices, analogous considerations hold). Since $p_{i,j}$'s 4-neighbourhood contains 3 lesional pixels, exactly one other box vertically or horizontally adjacent to $p_{i,j}$ is lesional. Assume that $j > 0$ and that this box contains $p_{i,j-1}$ (the other case is proved in the same fashion). The 4-neighbourhoods of $p_{i+1,j-1}$, $p_{i+1,j}$ and $p_{i+1,j+1}$ contain each 4 lesional pixels; if $i > 0$, the 4-neighbourhood of $p_{i-1,j}$ and $p_{i-1,j+1}$ contains at most 2 lesional pixels. Thus, none of these is a boundary pixel. $p_{i,j+1}$ is lesional and its 4-neighbourhood contains exactly 3 lesional pixels, so $p_{i,j+1}$ is a boundary pixel. We now prove that one and only one of $p_{i,j-1}$ and $p_{i-1,j-1}$ is a boundary pixel. If $i = 0$ or $p_{i-1,j-1}$ is a non-lesional pixel, then $p_{i,j-1}$ is a boundary pixel. Otherwise, if $i > 0$ and $p_{i-1,j-1}$ is lesional, then $p_{i,j-1}$ is not a boundary pixel, but $p_{i-1,j-1}$ is (since all the pixels in its 4-neighbourhood *but* $p_{i-1,j}$ are lesional). In either case, $p_{i,j}$'s 8-neighbourhood contains exactly 2 boundary pixels and our claim follows.

Consider now the second case – i.e. $p_{i,j}$ is neither a vertex nor the central pixel of B. For simplicity, we assume that $p_{i,j-1}$ is B's upper-left vertex, while $p_{i,j+1}$ is B's upper-right vertex. Then either $i = 0$ or $p_{i-1,j}$'s box is non-lesional. In both cases, $p_{i,j}$'s 8-neighbourhood contains 5 lesional pixels. $p_{i+1,j}$ is not a boundary pixel (since it is B's central pixel). Furthermore, exactly one between $p_{i,j-1}$ and $p_{i+1,j-1}$ is a boundary pixel: if $j > 1$ and $p_{i,j-2}$ is lesional then $p_{i,j-1}$ is a boundary pixel and $p_{i+1,j-1}$ is not; if $j = 1$ or $p_{i,j-2}$ is non-lesional then $p_{i+1,j-1}$ is a boundary pixel and $p_{i,j-1}$ is not (since its 4-neighbourhood contains only 2 lesional pixels). In the same way one can prove that exactly one between $p_{i,j+1}$ and $p_{i+1,j+1}$ is a boundary pixel. Then $p_{i,j}$'s 8-neighbourhood contains 2 boundary pixels, proving the theorem.

Chapter 4

Digital hair removal

An effective pre-processing phase may be crucial to improving the accuracy and the robustness of all of the algorithms used to process dermatoscopy images. The goal is to remove all of the spurious components that may interfere with the automated analysis, leading to a decisive improvement of accuracy and robustness.

During this phase, an algorithm for hair detection and removal is doubtless one of the most important. As mentioned in Chapter 3, our segmentation algorithm has a certain degree of robustness to the presence of hair. Besides that, correctly identifying the presence of hair can further boost the segmentation performance. Regarding the registration algorithm, which will be discussed in the next chapter, hair removal is a condicio sine qua non to meet the required target of accuracy. Generally speaking, most of the approaches for processing a lesion which requires an analysis of its border or its texture assume as source an hair-free image for optimal results. Although a few algorithms are partially robust to the presence of hair, an algorithm to effectively identify the hair presence could greatly improve the performance in any kind of processing performed on dermatoscopic images of melanocytic lesions.

The hair removal task can be coarsely split into two major steps. The first step is the hair segmentation, i.e. identifying all of the pixels in the image that belong to an hair. The second step is the production of an output image. Other additional steps such as hair thickness estimation (Subsection 4.2.5) or supplementary post-processing (Subsection 4.2.2) can be useful in certain contexts.

The first step is common to every approach proposed in the literature and it is widely independent from the processing that follows the hair removal phase.

The result of this phase is generally a mask that can be directly provided as an output or can be further processed by the second step. The different approaches proposed in the literature are presented in Subsection 4.1.1. In Subsection 4.2.1 we present the algorithm implemented in the Mole Mapper.

The second step strongly depends on the operation that will be performed on the image after the hair removal phase (e.g. a simple visualization, a segmentation, a registration with another image, a feature analysis, etc.). Regarding the following operation, its design, its goal and the input constraints should be taken into account to provide an output that minimizes the overall error. Generally, the most common approach is to provide a simple hair mask or to use an in-painting technique that tries to recover the lost information. In Subsection 4.1.2 some common in-painting techniques will be analyzed. In Subsection 4.2.4 we analyze the output provided by our algorithm for the segmentation and registration operations and for the visualization in different application contexts.

The accuracy of an hair removal software is crucial, since it affects the result of any other operation that relies on its output images. In Subsection 4.3.2 we compare our solution with DullRazor [56], which is widely used as a reference in hair removal algorithm comparison. Also, the computational time plays a big role since the hair removal is an essential component preceding most of the other processing operations. A comparison between DullRazor and Mole Mapper constraints is performed in Subsection 4.3.3 .

4.1 Related work

Many methods have been developed for digital hair removal in dermatoscopy images. They aim to produce an hair-free image to be used in a computer-aided detection (CAD) system. Each algorithm is generally composed of two main steps: (1) detecting the pixels that belongs to an hair; and (2) repairing the image by replacing the pixels with an estimated value to minimize the error of the subsequent lesions' analysis algorithms.

4.1.1 Hair pixels detection

The detection step attempts to identify both thin and thick hairs, avoiding the false positive due to the structures and lines that belong to the lesion pattern.

Frequently in the literature there is first a step to identify an initial hair pixels' mask, followed by its refinement.

For the first step a frequently used approach is a closing or closing-based top-hat morphological operators [57] [58]. Sometimes [59], [56] [60] multiple structuring elements are used for different directions. This approach tends to prefer the recognition of hairs with a certain direction, penalizing those less compatible with the structuring elements used. Kiani et al. [61] use Radon transform followed by a Prewitt filter with two different orientations. Niugen et al. [62] use a matched filter, cross-correlating the image with a series of 18 different rotations of a Kernel similar to an hair intensity profile. Some approaches rely on an edge detector as the first step. Toossi et al. [63] uses an adaptive Canny edge detector [64] whereas [65] and [66] use Steger's line detection algorithm [67]. Abbas et al. [68], [69], [70] use a 2-D derivative of a Gaussian filter that efficiently detects the lines in all directions.

The second step is generally a set of operations that try to enhance the accuracy of the initial mask. Each proposed algorithm generally presents different refinement steps that strongly depend on the weakest point of the initial mask, together with the properties of the specific dataset and the constraint that the specific hair removal algorithm wants to fulfill. An exhaustive dissertation on all of the methods that have been used is outside of the scope of this study. Briefly, the most frequently used techniques are: evaluating the geometrical parameters of the connected components to exclude non-hair structures, using morphological operators to fill small gaps, restoring the mask in the hair intersection, analyzing the relative position and shape of the components to try to fill longer gaps.

4.1.2 Hair pixels repair

Most of the algorithms proposed in the literature perform an image repair that aims at restoring the information occluded by the presence of hair. They aim to preserve the texture features and borders of the lesions, avoiding undesirable blurring or color-bleeding effects. Furthermore, the computational time required by some approaches is relevant. Frequently used approaches [69] for this step are linear interpolation, non-linear partial differential equation based diffusion, exemplar-based inpainting technique and fast marching algorithms.

Linear interpolation

A linear interpolation allows a simple and fast implementation at the cost of a poor performance in preserving features and texture. The basic idea is to replace the value of an hair pixel by averaging the values of nearby non-hairy pixels. A possible implementation is, for a given hair pixel $I(x, y)$, to find the non-hair pixels $I_1(x_1, y_1)$ and $I_2(x_2, y_2)$ closest to I which belongs to opposite sides of the hair segment. The new intensity value $I_n(x, y)$ can be computed as:

$$I_n(x,y) = I_2(x_2,y_2)\frac{d\left((x,y),(x_1,y_1)\right)}{d\left((x_2,y_2),(x_1,y_1)\right)} + I_1(x_1,y_1)\frac{d\left((x,y),(x_2,y_2)\right)}{d\left((x_2,y_2),(x_1,y_1)\right)}$$

where

$$d((a,b),(c,d)) = \sqrt{(c-a)^2 + (d-b)^2}$$

An implementation consistent with this model (or equivalent approaches) has been used in [56], [62] [58].

Non-linear PDE inpainting

A non-linear PDE diffusion based inpainting achieves generally better results compared to the linear interpolation technique at the cost of a higher computational complexity. Its main advantage is the capability to keep sharp boundaries, which is particularly useful when dealing with lesion borders. The main drawback is the introduction of blur during the diffusion process, which may negatively affect the lesion pattern. The procedure used for filling the holes in the image is inspired by the partial differential equation of the heat flow. The technique presented in the literature is generally an improvement and a refinement of the non-linear diffusion filter proposed by Perona and Malik. The pixel value is diffused according to [71]:

$$\frac{\partial u}{\partial t} = \nabla \cdot (c(c,y,t)\nabla u)$$

where $\nabla\cdot$ is the divergence operator, ∇ is the gradient operator and $c(c, y, t)$ is the diffusivity function. A PDE based approach has been used by [72] [57] [73] [59].

Exemplar-based inpainting

Exemplar-based inpainting methods [74] allow a better restoration of missing information, keeping the original image structure. They rely on texture synthesis algorithms, which are useful for generating a large image region using sample textures, combining them with non-linear PDE diffusion. This approach combines the advantages of both techniques: filling holes, keeping a consistent texture and respecting at the same time the constraint imposed by the surrounding linear elements. The main drawbacks of this approach are the computational complexity and a set of parameters (e.g. the processing window size and the number of iterations) that are hard to determine a priori. In the literature exemplar-based inpainting methods have been used in [66] [45] [68].

Fast marching technique

The fast marching technique solves the main problems of exemplar-based inpainting approaches, since it is parameter-less and less computationally intensive. This inpainting method has been proposed by Bornemann and März [75]. approach, depending on the measure of the coherence strength, the inpainting procedure conveniently switches between diffusion and directional transport. As opposed to the PDE and exemplar-based inpainting, it is not an iterative technique; this allows for implementation at least an order of magnitude faster compared with PDE and exemplar-based methods, as stated in [75]

For repairing pixels in dermatoscopic images, the fast marching technique was utilized in [70], [69], [76].

4.2 Proposed algorithm

The proposed algorithm is composed of 3 different steps. In the first step (4.2.1) a mask is produced using the information coming from two different sets of DoG filters. In the second step (Subsection 4.2.2), the mask is processed, removing small speckles and adding missing components with an hair shape. An optional post-processing step is presented in Subsection 4.2.3. In the third step (4.2.4)the output is produced. The output can be a simple mask, or an inpainting algorithm can be applied, depending on the requirements. All the steps requires the information about the average hair thickness in the processed picture. In Mole

Mapper, the picture resolution and zoom factor are fixed, so there is an a priori knowledge of this information. To obtain such a information in a different dataset we present in (4.2.5) a technique for the estimation of average hair thickness.

4.2.1 Extracting the hair mask

The mask extraction is performed using a set of DoG (i.e. Difference of Gaussians) filters. It performs the extraction of a *loose mask* which contains most of the true positive together with some false positive, and a *strict mask* that contains almost only true positive. The *strict mask* is then used to validate the loose mask connected components. A speckle remover is then applied. Finally, some non-validated components of the loose mask are added back to the final mask, depending on their shape.

In the following paragraphs, an overview of the DoG Filter is given, followed by a detailed explanation of each algorithm step.

DoG Filter

A DoG filter is an operator used to enhance the features in an image. It is computed as the difference between two Gaussian Filters with different standard deviation. The general formula of a DoG filter Γ applied to the image I is:

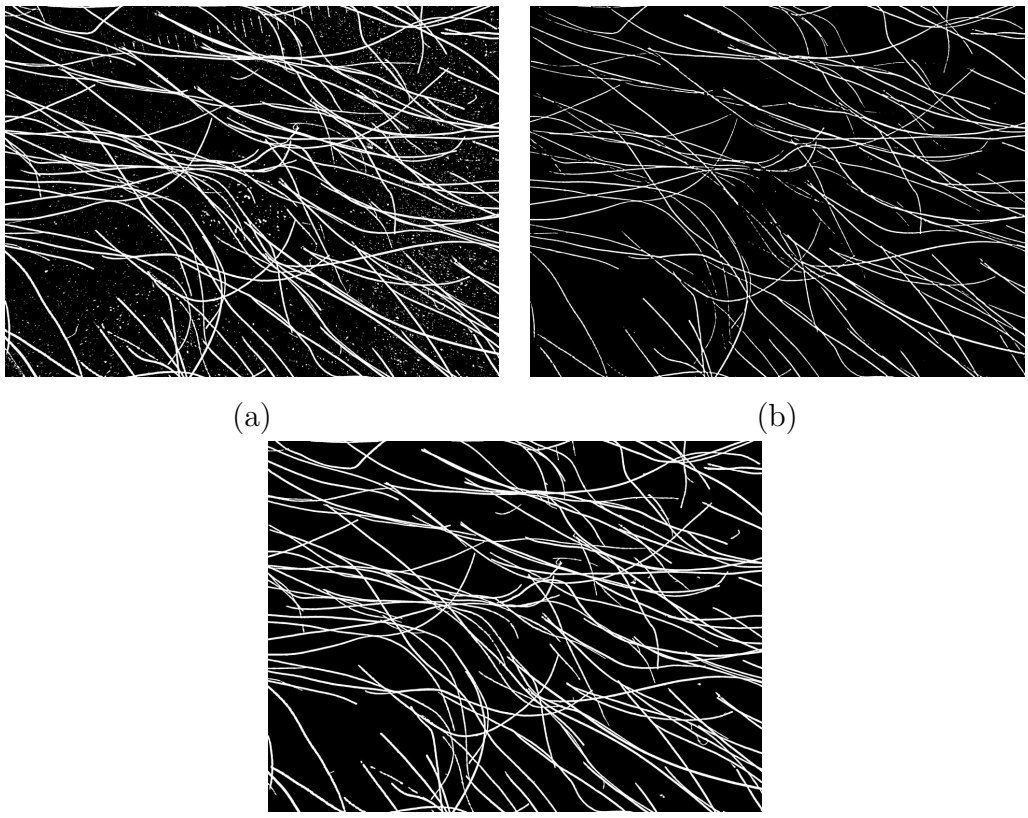
$$\Gamma_{\sigma,K\sigma}(x,y) = I * \left(\frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/(2\sigma^2)} - \frac{1}{2\pi K^2\sigma^2} e^{-(x^2+y^2)/(2K^2\sigma^2)} \right)$$

where :

σ	is the standard deviation of the smaller Gaussian Filter
$K\sigma$	with $K > 1$ is the standard deviation of the bigger Gaussian Filter
$*$	is the convolution operator

Loose and Tight masks extraction

The goal of this phase is to obtain a *loose mask* that contains with high probability all the pixels that belong to an hair and a *strict mask* that identifies with a very high confidence pixels that belong to an hair.



(a) (b)

(c)

Figure 4.1: An example of *loose mask* (a) *tight mask* (b) and their combination (c).

Both masks are computed using some DoG filters $\Gamma_{\sigma_1,\sigma_2}$. The use of this operator, which relies on a circular structuring element, led to a notably better result on our image-set compared with the more common top hat with linear structuring element approach.

The standard deviation parameters of the filters (σ_1, σ_2) are functions of the average hair thickness expressed in pixels h. In Mole Mapper that value can be considered constant, but for a general use case, a method for its estimation is shown in Subsection 4.2.5.

Each DoG operator is followed by a thresholding phase obtaining a binary mask that can be merged together with the masks obtained with different standard deviation and thresholding parameters.

The tight mask is defined as:

$$M_{\text{tight}} = \Gamma_{\frac{h}{\sqrt{3}},\sqrt{3}h} > t_1$$

; this allows a solid recognition of blobs of a size comparable to h, regardless of their effective shape.

The loose mask is defined as:

$$M_{\text{loose}} = \Gamma_{\frac{h}{\sqrt{3}},\sqrt{3}h} > t_2 \;||\; \Gamma_{2,\sqrt{3}h} > t_3 \text{ with } t_2 < t_1$$

the first element of this mask is the same DoG filter used in the tight mask but with a looser threshold, allowing the selection of hair segments in a low contrast zone. In addition, the second term allows the selection of blobs that are thinner than an average hair. This is useful to include the extremes of an hair, or very thin hairs.

In our dataset we obtained a good result using $t_1 = 8,\ t_2 = 2,\ t_3 = 4$

Combining the two masks

At this point we have the mask M_{tight} which contains pixels that are most likely a true positive and the mask M_{loose} which suggests pixels that belong to an hair with a weaker confidence. We merge the two sets of information in a resulting mask M_{merged} whereby each pixel is defined as:

$$M_{\text{merged}}(j,i) = \begin{cases} 1 & \text{if } M_{\text{loose}}(j,i) == 1 \text{ AND } \exists\, i_2, j_2 \; s.t. \; (j_2, i_2) \in C_{M_{\text{loose}}}(j,i) \\ 0 & \text{otherwise} \end{cases}$$

where $C_M(j,i)$ is the set of all the pixels (i_2, j_2) that belong to the same connected component of (j,i) in the image M. In other words, we are selecting all the connected components in M_{loose} with at least one pixel set in the M_{tight} mask.

4.2.2 Mask post-processing

On the mask M_{merged} generated in the previous step a post-processing phase is performed. In particular, some smaller area needs to be removed and some additional connected components of M_{loose} should be included in the M_{merged} mask.

For removing the smaller spurious areas, all of the connected components in M_{merged} are enumerated. For each connected component the bounding box is

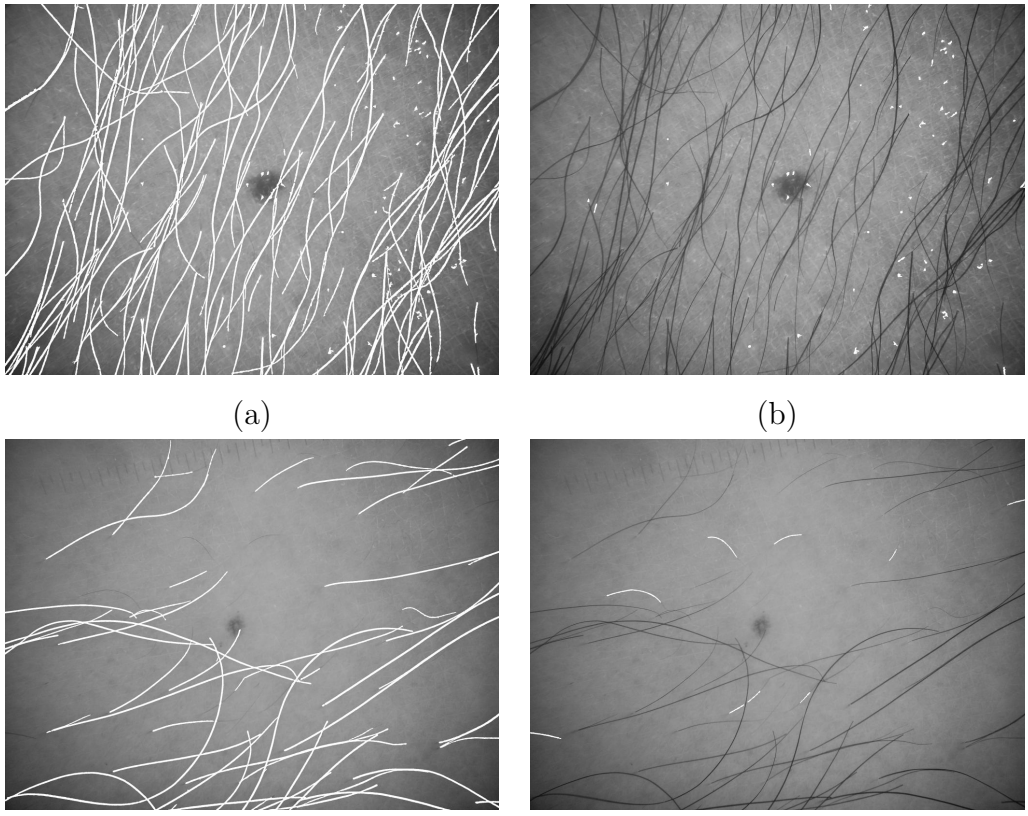


Figure 4.2: Hair removal mask post-processing. From the initial mask (a) the small spurious area selected for removal (b). From another initial mask (c) small components are added, depending on their shape (d).

detected and its width W_b and height H_b are computed. All of the connected components such that:

$$H_b + W_b < \alpha h$$

are then discarded.

For adding to the mask M_{merged} the connected components improperly ignored during the masks merging phase, all of the connected components in M_{loose} are first enumerated. Then for each component the area A_b and the perimeter P_b are computed. All the connected components such that:

$$H_b + W_b > \alpha h \text{ AND } \frac{A_b}{P_b} < \beta$$

are then added to the mask M_{merged} .

In our dataset we obtained a good result using $\alpha = 15, \beta = 10$

4.2.3 Optional: Hair graph filling

For some applications it may be useful to identify all the disconnected hair segments in the mask and perform an hair fitting operation to reconnect them. We developed an algorithm to perform this task which is not run by default during any digital shaving process but can be optionally activated by the user. The algorithm is composed of three steps: first, a thinning algorithm is applied to the entire mask, then the parameters of the hair extremes are computed and finally a compatibility parameter between each pair of extremes is computed and the gap between the two extremes is filled where the compatibility is verified. Now each step will be analyzed in more detail.

In the first step the skeleton of the hair mask is generated. The algorithm proposed in [77] has been chosen for its connectivity preserving properties and has been implemented.

In the second step the skeletonized mask is scanned to identify the position of the hair endpoints. We define the set of all endpoints as $P = \{p_1, p_2 ... p_n\}$. For each $p_i \in P$ the angle ϕ_i is computed, performing a line fitting using the p_i neighbors in the skeletonized mask. More formally, given the endpoint p_i, which belongs to the connected component C_j, we defined the set of his neighbors N_i as the set of points in C_j whose distance from p_i is less than $4h$, where h is the average hair thickness. The PCA is performed on the coordinates of points belonging to N_i to retrieve the prevailing direction ϕ_i.

In the last step for each pair of endpoints $(p_i, p_j), i \neq j$ a compatibility is evaluated depending on their distance and their direction. We consider a pair compatible if:

$$d(p_i, p_j) < 50h$$

and

$$d_\theta(\text{atan2}((p_i^y - p_j^y, p_i^x - p_j^x), \theta_i) < \theta_\epsilon$$
$$d_\theta(\text{atan2}((p_j^y - p_i^y, p_j^x - p_i^x), \theta_j) < \theta_\epsilon$$

with:

$$\text{atan2}(y,x) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & x > 0 \\ \tan^{-1}\left(\frac{y}{x}\right) + \pi & y \geq 0, x < 0 \\ \tan^{-1}\left(\frac{y}{x}\right) - \pi & y < 0, x < 0 \\ +\frac{\pi}{2} & y > 0, x = 0 \\ -\frac{\pi}{2} & y < 0, x = 0 \\ \text{undefined} & y = 0, x = 0 \end{cases}$$

where $d(\cdot,\cdot)$ is the Euclidean distance, h is the average hair thickness, $d_\theta(\cdot,\cdot)$ is the minimum absolute difference between two angles, p^x and p^y are the x and y coordinates of the point p and θ_ϵ is a parameter for the maximum angular difference ($\theta_\epsilon = \frac{\pi}{20}$ in our experiments).

For each compatible pair the corresponding hair segments are then connected, performing an interpolation between the two end-points.

4.2.4 Output generation and Inpainting

The output provided by the digital shaving process depends on what will be the usage of such a output in the following steps. In Mole Mapper a digitally shaved image can be used for visualization or as the input of the other image processing algorithms.

Regarding the visualization, the main usage of a shaved image is for a single lesion evaluation or a side-by-side lesion comparison. In this case we decided to allow a full control on the shaving output by the user. It is possible to hide hairs with a mask, perform a bilinear interpolation inpainting or apply the Telea [78] fast marching algorithm. Visualization of shaved images is also performed for small previews in the "lesion overview" screen. In this case, due to the timing constraints and the usage of the shaved images (it consists only of taking a quick look over a big set of shaved images) the bilinear interpolation has been chosen.

In the case that the output is used by other processing algorithms, we always provide the hair segmentation mask, without any other inpainting technique. The main reason for avoiding inpainting is that on one hand a bilinear interpolation can degrade the information needed for the automatic analysis, and on the other

hand the computational cost of the algorithms that preserve lesion features is too high to extensively apply it to all of the acquired full resolution images. For these reasons all the lesion analysis algorithms have been designed and implemented for handling a *don't care* option defined at pixel level by a binary mask. This implies a slightly higher complexity of the algorithm for handling the occlusion, but at the same time this avoids the propagation of wrong results due to incorrect information recovery during the digital shaving phase.

4.2.5 Optional: Average hair thickness estimation

The hair removal algorithms presented in the literature frequently rely on structuring elements or morphological operations with a pre-defined size. Using these algorithms on images with a different resolution or magnification can lead to a consistent drop of robustness and sometimes to an unexpected behavior. For this reason our Hair Removal algorithm has been developed using the average hair thickness h (in pixels) as its parameter. It's sufficient to perform the estimation of its value only after installing the application on a device and if the pair camera/optic was not tested before. The parameter estimation may also be performed when testing the algorithms on a dataset not acquired with a Mole Mapper device.

Our algorithm for thickness estimation performs a coarse hair segmentation as described in Subsection 4.2.1, varying the h parameter in the set $\{1, 4, 16\}$. Each mask is analyzed by randomly selecting some areas and verifying the consistency with the hair shape model. For each mask that passes the consistency checking a morphological erosion operation with a 3×3 element is iteratively applied. The parameter h is then defined as the number of iterations necessary to reduce the mask of 90%. The h evaluated in the different masks are finally averaged to compute the final h value.

4.3 Experimental evaluation

Subsection 4.3.1 briefly describes the experimental setup. Subsection 4.3.2 defines the comparison metric and evaluates the accuracy of our algorithm. Finally, in Subsection 4.3.3 the computational time is evaluated on a PC and a mobile device.

4.3.1 Experimental Setup

The entire dataset (referred to as *Full-Set*) is composed of 35 3264 × 2448 images acquired with an Asus Tranformer TF201 equipped with a DermLite FOTO dermatoscope. The set was acquired from different patients and different body areas to well represent images with greatly different hair density. For a subset of 10 images (referred to as *Subset-10*) a ground truth has been provided. In this subset, images with an extensive presence of hair and images with few and tiny hairs are both present.

We compared our algorithm to DullRazor [56] software available at `http://www.dermweb.com/dull_razor/` which is widely used in the field of dermatology.

The ground truth was produced by a human operator who manually marked pixels of the image as hairs. After this first round, for very hairy images we noticed the presence of some false negatives (i.e. some pixels that belonged to hair but were marked as skin). This is due to the fact that the manual segmentation on hairy images performed by a human operator is a very demanding task in terms of time and focus, so the probability of missing some small hairs during the labeling process is not negligible. To solve this problem, we performed an automatic analysis with both DullRazor and our algorithm on the dataset for producing a mask of the probable false negative. In a second round, another human operator reviewed the original ground truth mask, taking into account the suggestion coming from the automatic analysis. Finally, in a third round all masks were checked again by a third different human operator to check the results. All of the masks passed this stage without any error being identified.

4.3.2 Accuracy

The accuracy test was performed in two different phases. First, we compared our algorithm to DullRazor using a ground truth on *Subset-10*, then we analyzed our algorithm and DullRazor's behavior on the *Full-Set* without a ground truth to study the trend of false positives and false negatives.

Comparison Metric

For the quantitative evaluation we decided to avoid a comparison on a pixel-by-pixel basis. The reason comes from the usual hair shape: the connected

(a) (b) (c)

Figure 4.3: Hair removal performed by DullRazor and our algorithm. (a) Original image. (b) DullRazor. (c) Our algorithm hair removal and repair.

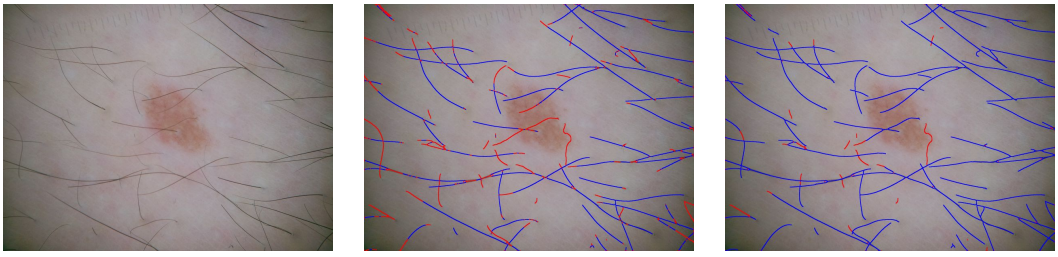
components in the mask that identify hairs have an amount of pixel belonging to its frontier (i.e. the perimeter) which is comparable to the total amount of pixels belonging to the entire region (i.e. the area). This implies that an error of a few pixels in estimating the hair thickness along the border can lead to a very high error rate even if the entire hair has been correctly identified. This is particularly harmful when this kind of error exceeds the error due to the wrong evaluation of a whole hair group or big hair segments.

The quantitative evaluation of the error of the mask T under testing is performed by measuring what fraction of hair length has been incorrectly classified. The divergence d from the ground truth mask R is evaluated using a XOR metric:

$$d = \frac{\mathrm{FP} + \mathrm{FN}}{\mathrm{TP} + \mathrm{FN}}$$

where FP (false positive) is the length of the segments in T without a correspondence in R, FN (false negative) is the length of the segments in R without a correspondence in T, TP (true positive) is the length of the segments in R with a correspondence in T.

More precisely, to perform the comparison, we compute the skeletons T_s, R_s of T and R respectively, using the Guo and Hall thinning algorithm [77]. The T_s skeleton is then registered on R_s. The registration is performed by putting



(a) (b) (c)

Figure 4.4: Comparison of DullRazor (b) and our algorithm (c) with the ground truth using the XOR metric on the skeletons. True Positive are in blu; False Negative are in Red; There are no area classified ad False Positive.

in relation each pixel of T_s with all the pixels in R_s having the distance below the average hair thickness h. The same registration process is performed again switching the roles of T_s and R_s. This produces an N-to-M relationship graph between the pixels of the two skeletons that are used to identify whether a pixel in a mask has one or more correspondences in the other mask. This behavior is good for the inner skeleton points, but may be a too relaxed a condition near the end-point area. This problem is fixed by performing a convenient pruning of the relationship graph in the proximity of the endpoints. After this process we define as TP the amount of skeleton pixels in R_s that appear in a relationship, FN the amount of skeleton pixels in R_s that do not appear in any relationship, and FP the amount of skeleton pixels in T_s that do not appear in any relationship.

Comparison with a ground truth

Using the comparison metric just defined we measured the accuracy of our algorithm and that of DullRazor using the segmentation performed by the human operator as ground truth. The test performed on *Subset-10* shows (see Table 4.1) an average divergence of 15.9% for our algorithm compared to 41.7% for DullRazor. It is interesting to notice that practically the entire error is due to the presence of false negatives (15.7% and 41.6%, respectively) i.e. the main problem is that both algorithms are unable to recognize some hairs.

	d (avg)	d (std)	$\frac{\text{FP}}{\text{TP+FN}}$	$\frac{\text{FN}}{\text{TP+FN}}$
DullRazor	41.7%	18.8%	0.1%	41.6%
Our Algorithm	15.9%	14.58%	0.2%	15.7%

Table 4.1: Divergence d_s on the *Full-Set* (average, standard deviation, false positive rate, false negative rate) of DullRazor and our algorithm using the ground truth provided by the human operator.

DullRazor Direct Comparison

In light of the extremely low false positive rate highlighted in the ground truth comparison result, and taking into account the very high cost of producing a ground truth, we decided to perform a comparison on the *Full-Set* comparing our algorithm with DullRazor directly.

In this test we adopted DullRazor as a ground truth, using the same comparison metric defined before. The average divergence is 38.0% with 0.8% of false negative rate and 37.2% of false positive rate.

Performing the same comparison only on *Subset-10* used in the previous experiment with the ground truth, we obtained an average divergence of 35.3% with 0.6% of false negative rate and 34.7% of false positive rate.

Summing up the two results, we observed on a small dataset that both algorithms have a negligible false positive rate and our algorithm is able to identify 38.0% more hair without a negative impact on the false positive value. The comparison on the entire dataset seems to confirm this trend, but a more extensive test using a ground truth may be required for a more accurate evaluation.

Dataset	d (avg)	d (std)	$\frac{\text{FP}}{\text{TP+FN}}$	$\frac{\text{FN}}{\text{TP+FN}}$
Subset-10	35.3%	14.36%	0.6%	34.7%
Full-Set	38.0%	12.82%	0.8%	37.2%

Table 4.2: Divergence d_s on *Full-Set* and *Subset-10* (average, standard deviation, false positive rate, false negative rate) of our algorithm using DullRazor as ground truth.

4.3.3 Computational resources

We measured the execution time of our digital hair removal algorithm on an X86 Windows PC and we compared it with DullRazor. Also, the execution time on an Android ARM device has been measured to check the fulfillment of the Mole Mapper user interaction timing constraints.

On the X86 platform the test was performed on the *Full-Set*, measuring the execution time for hair mask generation and the bilinear inpainting. The PC was equipped with an Intel i7-4700MQ CPU with 16GB DDR3 1600MHz, running Windows 8.1 64-bit. The time spent for image reading and decoding was not taken into account. The DullRazor execution time was tested on the same machine using the binary for Windows publicly available. Since in this case a correct timing excluding the I/O was not possible, the input and output operations were performed on a RamDisk to reduce their repercussion on the overall execution time.

The ARM device was a LG Nexus 5 equipped with a Qualcomm Snapdragon 800 with 2 GB of LPDDR3-1600 RAM, running Android 4.4.2.

Our algorithm execution times (see Table 4.3) were almost 30 times lower than DullRazor on the X86 platform. The hair mask production of our implementation requires less than a second on the Desktop PC and about 5 seconds on a mobile device using full resolution images.

On the actual Mole Mapper implementation our shaving algorithms is not run on the entire image but only on a rectangular ROI that surrounds the lesion. The ROI is obtained using a preliminary coarse segmentation on a scaled-down version of the image. Performing the computation only inside the ROI allowed us to reduce the average computational time to less than a second, which leads to a tolerable delay for the end user.

4.4 Conclusions

We developed an algorithm for hair removal from dermatoscopic images composed of a detection phase and an optional repair phase.

The detection is performed by merging the information from two different masks obtained with DoG filters. A post-processing phase then removes small speckles and includes additional hair-shaped components missed in the previous

	PC X86	Android ARM
DullRazor Hair mask	25718	-
Our Algorithm Hair Mask	952	5347
Our Algorithm Inpainting	38	235

Table 4.3: Execution time (ms) on a PC i7-4700MQ and an Android Nexus 5 of DullRazor and our algorithm.

step. An optional step of connected component analysis allows the restoration of disconnected hair segments.

The optional repair phase is performed only for visualization purposes, whereas only an occlusion mask is provided to the other image processing algorithms. Bilinear interpolation and a Telea fast marching inpainting were implemented and can be used depending on time constraints and the user's needs.

Our algorithm proved to be very accurate when compared to DullRazor, with an error rate that was almost three times lower. We also verified that substantially all of the error came from the presence of false negatives, where most of the missing pixels resulted from the presence of very thin hair, a defect that does not actually affect most of the algorithms that require Virtual Shave as a preprocessing module. Conversely, the rate of false positives is practically negligible.

The algorithm is also very fast, almost 30 times faster than DullRazor. The proposed technique of merging the information coming from both a tight and a loose masks reduces the need for major adjustments in the refinement phase. It avoids complex post-processing operations and the examination and comparison of a high number of connected components, keeping the overall computation time comparable with the execution of a few morphological filters.

Chapter 5

Dermatoscopic images registration

Performing the registration of dermatoscopic images is crucial to studying the evolution of a lesion over time. It helps to simplify the comparisons performed by both automated algorithms and human operators.

Even though the general problem of registering two images is well treated in the literature, very few works addresses the problem for skin lesions in dermatoscopic images. An analysis of the current state of the art is performed in Section 5.1.

Due to the lack of standard approaches and a rigorous analysis of the problem, in Section 5.2 we characterize the problem, analyzing the relationship between the lesion evolution and the deformation model, highlighting at the same time the challenging aspects and the constraints.

In Section 5.3 we propose a registration algorithm that relies on the robustness of MEDS segmentation. It tries to identify a good match between the segmentation performed on two different pictures. If a good match is not detected, two more precise and computational intensive phases will be performed. This algorithm, actually implemented on Mole Mapper, allows a very good overall accuracy, while at the same time keeping the average computational cost very low.

The performance of the proposed algorithm will be analyzed in Section 5.4. The accuracy and computational time needed for the entire algorithm and for each individual step is evaluated. The overall accuracy is finally compared with

the accuracy of registration performed by a human operator. The computational time is finally analyzed and compared with the constraints imposed by the user interface and the workflow.

5.1 Related work

There are few methods proposed in the literature for the registration of melanocytic lesions in dermatoscopical images. The review in this section is then extended to the segmentation of different kinds of skin lesions (e.g. psoriasis, hamangiomas etc.).

Pavlopoulos [79] proposed a hybrid stochastic-deterministic method for the registration of malignant melanoma images. The scaling and rotation parameters are determined using a log-polar transformation technique, whereas the two translation parameters are obtained using a hill-climbing optimization scheme.

Maletti et al. [80] registered digital images of the lesions of psoriasis. They assume that the shape and size of the portion of the skin to be tracked is constant across different images. They first perform a rigid alignment assuming equivalence in location correspondence and afterward they apply a combined contextual registration and alignment scheme.

Delgado et al. [81] proposed an algorithm for registering psoriasis images. A segmentation of the lesion is performed and then the rotation and translation parameters are estimated using the statistical area moments.

Zambanini et al. [82] presented a method for registering hemangioma images. They used scale-invariant feature transforms (SIFT) [83] to find interest points inside the hemangioma area, and then the homography between the two images is estimated by means of Random Sample Consensus (RANSAC) [84].

Anagnostopoulos et al. [85] proposed a registration method of melanocytic lesions from digital dermoscopy. A modified version of the SIFT algorithm is used followed by the computation of the affine transformation parameters using RANSAC. The modified SIFT algorithm is called ROI-SIFT. It consists of a first run of the SIFT algorithm with "hard" parameters that produce a low number of feature points. Then, according to the position, the scale and the orientation of the obtained features, an expanded Region of Interest (ROI) is defined. Finally, a second run of the SIFT algorithm with "soft" parameters is performed keeping

only the keypoints that belong to the ROIs.

5.2 Major issues and constraints

The registration of dermatoscopic images is necessary to identify the differences between the lesions in two different photo sessions. However, frequently the differences are not imputable to a lesion evolution but to the different conditions during the acquisition process. In Subsection (5.2.1) most common variations introduced by the acquisition process are listed. The requirement for robustness to these variations, while retaining the ability of highlighting the lesion evolution, strongly affects the choice of the transformation model (see Subsection 5.2.2)

5.2.1 Variations different from evolution

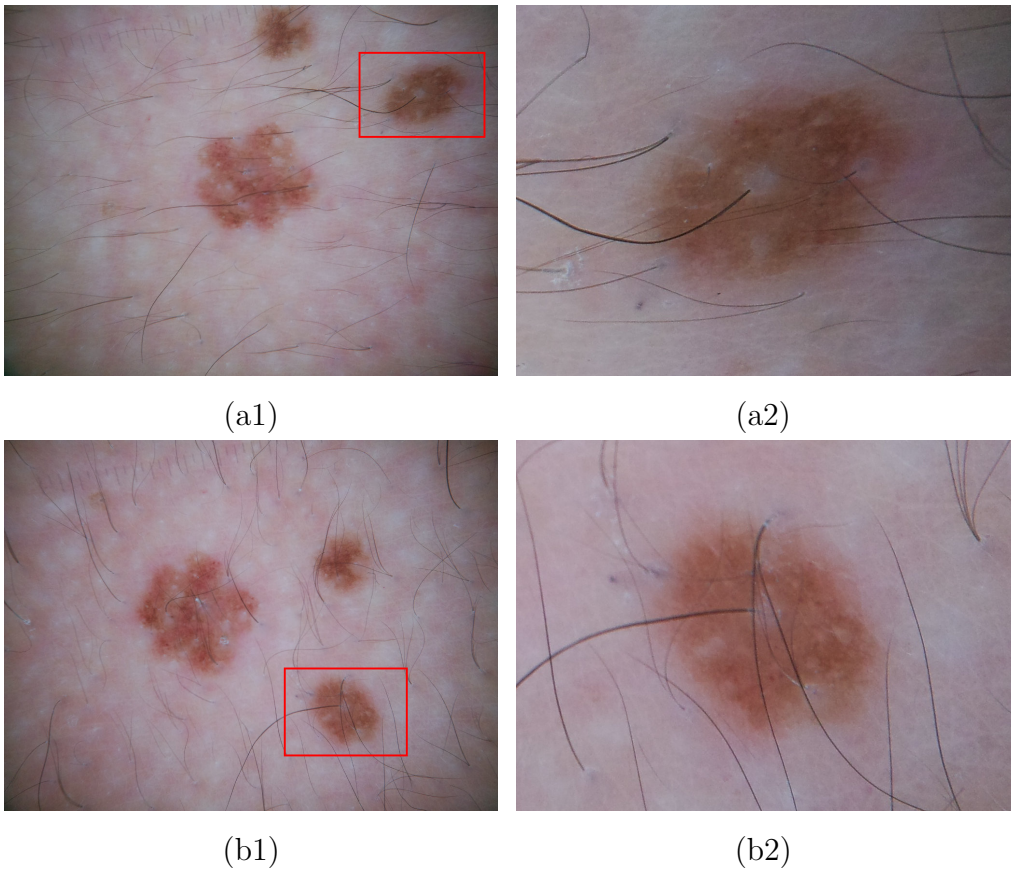
Different Light conditions

Figure 5.1: Two pictures of the same lesion with notable light changes.

Even though the image acquisition in Mole Mapper is performed with a specific illumination provided by the dermatoscope, we noticed a certain degree of variability to the illumination of the acquired images. In our tests we identified three different main causes for that. The first is the lack of manual exposure settings which can lead to different values being computed by the auto exposure logic. The second cause is the variability of the led light depending on the temperature and the battery provided current. The last cause is the environmental light: although in most cases the light provided by the dermatoscope was orders

of magnitude greater than the ambient light, we noticed some variations in the captured images in the presence of strong environmental lights.

Non-uniform illumination



(a1) (a2)

(b1) (b2)

Figure 5.2: Example of non-uniform illumination. Different position of a lesion inside two dermatoscopic images with similar illumination (a1,b1) can lead to strong differences when considering a ROI around the lesion itself (a2,b2).

The illumination produced by the dermatoscope may be non-uniform across the entire image surface. Generally, the intensity is higher in the middle and lower in the area close to the image border. This issue is particularly noticeable for lesions with a greater area, which necessarily take up a bigger portion of the picture. The problem is imperceptible for smaller lesions and if attention is paid to keeping its position in the center of the image.

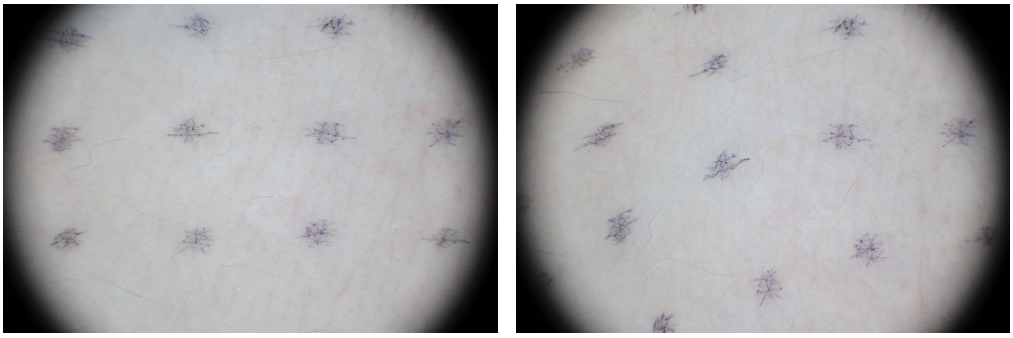


Figure 5.3: Deformation produced by the pressure of dermatoscope on the skin.

Skin Deformation

The pressure of the dermatoscope on the skin can produce a significant deformation of the skin itself. This problem marginally affects almost every captured image, but in certain circumstances its influence can be massive. The overall deformation can be particularly severe when the relative position between the lesion and the dermatoscope is adjusted without separating the lens from the skin. Avoiding this behavior during the acquisition phase greatly reduces the problem, although a marginal effect remains.

Nodular Naevi

The acquisition process performed by our device always requires a prefect adherence between the skin and the lens of the dermatoscope. This implies that prominent components on the skin are flattened under the pressure of the lens. In the case of nodular nevi, this flattening process can produce dramatically different results in different capture sessions. Sometimes the variation is so relevant that an expert dermatologist is not able to identify whether two images are depicting the same lesion. For this reason we decided to completely remove the nodular nevi in our evolution evaluation process. This has no major relevance from a clinical point of view, since a suspicious nodular nevus is generally excised promptly without tracking its evolution in the next visits.

Blurry images

This error sometimes occurs in our dataset; it is generally due to a human error in the acquisition phase. Frequently it is caused by setting the wrong dioptric

correction value in the dermatoscope, by having a gap between the lens and the skin in the lesion area and by quickly moving the device during the acquisition, causing a motion blur effect.

Different surrounding elements

Figure 5.4: Many elements (highlighted in red) surrounding the melanocytic lesion that will be absent in the other images of the same lesion.

There are many elements that belong to the surrounding area of the lesion and can be different in images acquired in two different sessions. This problem does not generally affect most of the evaluation algorithms but can be a weak point for algorithms that take into account textural elements or features of the surrounding skin. Some tedious major changes include different average skin color, different hair length or the presence of a temporary imperfection (e.g. air bubble, dust etc.). It is important to highlight that in general the occurrence of imperfection over the lesion is much lower than over the surrounding skin. This is because a picture with an evident imperfection over the lesion area will most likely be discarded and another one will be acquired by the dermatologist.

5.2.2 Transformation model

Choosing the best transformation model is crucial to obtain an adequate accuracy together with an acceptable execution time and satisfactory robustness to the variations. The presence of skin deformation suggests the use of local *elastic*

or *non-rigid* transformations. These models require an efficient and robust alignment of local lesion structures, but the weak presence of features and structures on the lesion's surface make this choice hardly viable. In addition, the lack of an operative model on the melanocytic lesion's evolution makes it difficult to estimate whether local movements or resizements are related to a skin deformation or to the evolution of the lesion itself. For these reasons we decided to use a global transformation model. This choice is also consistent with the transformation model used in the literature and presented in Section 5.1.

We additionally decided to exclude the estimation of scale since the focal length of the Mole Mapper is fixed.

Taking into account all of the previous considerations, we converged to a roto-translation model that requires the estimation of 3 independent parameters: a pair $\mathbf{T} = (T_x, T_y)$ that represent the translation and a value θ for the rotation. Using a homogeneous coordinate system, we can express the transformation with the following matrix multiplication:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & T_x \\ -\sin\theta & \cos\theta & T_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

where $[x, y, 1]^T$ and $[x', y', 1]^T$ are the coordinates before and after the registration, respectively.

5.3 Proposed algorithm

The proposed algorithm is composed of three different trials (Subsection 5.3.2) executed sequentially. Each trial attempts to estimate the best values for the parameters $\mathbf{T}$ and θ of the transformation model for a pair of dermatoscopic images named reference and test. The reference image is kept virtually still, whereas the test needs to be transformed to find the best alignment. The three trials have increasing effectiveness at the cost of a longer execution time. There are, in order, a *single border* comparison, followed by a *multi-border* comparison and finally a *bruteforce border* comparison approach. If a test finds a match with sufficient confidence, the execution is stopped and the following trials are not performed. All the trials use the same primitives that will be presented in the following paragraphs (Subsection 5.3.1).

5.3.1 Shared primitives

Border extraction

The border extraction is a primitive that returns a set of borders for different thresholds on the lesion image. For a given threshold the resulting border is computed very close in spirit to our segmentation algorithm MEDS with some minor differences. The following are the border computation steps:

- The image is converted to Grayscale, contrarily to the segmentation, the PCA is not performed and simply the blue channel is used. This allows a faster computation while keeping a good result at the same time.
- A thresholding operation is performed using the threshold value specified as input. The result is a binary mask.
- A post-processing is performed to keep only a single chain of pixels which represents the output border B. The algorithm used is the same as that used for Melanocytic Lesion Segmentation (Chapter 3, Subsection 3.2.5).

Border smoothing

This primitive is used to remove the high frequency component of the borders. The smoothing process is close in spirit to [86] We first express the border b in terms of two functions:

$$b = \{x(t), y(t)\} \;\; t \in [0, 1] \tag{5.1}$$

$x(t)$ and $y(t)$ are then convolved with a Gaussian kernel g:

$$g(t, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-t^2}{2\sigma^2}} \tag{5.2}$$

obtaining:

$$\begin{aligned} X(t,\sigma) &= x(t) * g(t,\sigma) &= \int_{-\infty}^{+\infty} x(u) \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-u)^2}{2\sigma^2}} du \\ Y(t,\sigma) &= y(t) * g(t,\sigma) &= \int_{-\infty}^{+\infty} y(u) \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(t-u)^2}{2\sigma^2}} du \end{aligned} \tag{5.3}$$

Since the border is a closed curve, $x(t)$ and $y(t)$ are treated as periodic functions during the convolution, eliminating the edge related effects. The resulting smoothed border B is finally:

$$B = \{X(t), Y(t)\} \;\; t \in [0, 1] \tag{5.4}$$

Border matching

This primitive finds the best correspondence of a tested border B_T with a reference border B_R. This is performed in three steps: First the centers of mass $\mathbf{C}_R, \mathbf{C}_T$ of B_R, B_T respectively are computed to find the translation parameter $\mathbf{T} = (T_x, T_y)$, then the polar coordinates P_R, P_T are computed around the centers of mass; finally, the best value for the translation $\bar{t}$ that minimizes the difference between P_R and P_T is estimated to obtain the rotation parameter θ.

The physical definition of the coordinates $\mathbf{C}$ of the center of mass given a system of particles $\mathbf{P_i}$, $i = 1, .., n$, each with mass m_i that are located in space with coordinates $\mathbf{r}_i$, $i = 1, .., n$ is:

$$\mathbf{C} = \frac{1}{M} \sum_{i=1}^{n} m_i \mathbf{r}_i \tag{5.5}$$

A border B can be interpreted as a finite set of discrete particles with unitary mass ($m_i = 1$) and coordinates $\mathbf{p_i} = (x_i, y_i)$ in a 2D space. The coordinates $\mathbf{C} = (C_x, C_y)$ can be computed as:

$$C_x = \sum_{i=1}^{n} \frac{1}{n}(x_i) \qquad\qquad C_y = \sum_{i=1}^{n} \frac{1}{n}(y_i) \tag{5.6}$$

The computation of the polar coordinates for a border B is performed using its center of mass $\mathbf{C}$ as the origin point. Each pair of coordinates $\mathbf{p_i} = (x_i, y_i)$ is

transformed into a pair $\mathbf{q_i} = (r_i, \rho_i)$ using the following formula:

$$\begin{aligned} r_i &= \sqrt{(x_i - C_x)^2 + (y_i - C_y)^2} \\ \rho_i &= \text{atan2}(y_i - C_y, x_i - C_x) \end{aligned} \tag{5.7}$$

with:

$$\text{atan2}(y,x) = \begin{cases} \tan^{-1}\left(\frac{y}{x}\right) & x > 0 \\ \tan^{-1}\left(\frac{y}{x}\right) + \pi & y \geq 0, x < 0 \\ \tan^{-1}\left(\frac{y}{x}\right) - \pi & y < 0, x < 0 \\ +\frac{\pi}{2} & y > 0, x = 0 \\ -\frac{\pi}{2} & y < 0, x = 0 \\ \text{undefined} & y = 0, x = 0 \end{cases} \tag{5.8}$$

To allow for easy comparison between two different borders, the information of polar coordinates is summarized into a bucket data structure H. Roughly speaking, the interval $(0, 2\pi]$ is partitioned into m different sub-intervals ($m = 256$ in our tests). Each sub-interval is related to a circular sector with a central angle of size $2\pi/m$. The radii r_i of all the coordinates that belong to this circular section are averaged and the result is assigned to $H(j)$. More formally:

$$Q_j = \left\{ q_i = (r_i, \rho_i) | \rho_i \in \left(\frac{2\pi j}{m}, \frac{2\pi(j+1)}{m} \right] \right\} \quad j = 0, .., (m-1) \tag{5.9}$$

$$H(j) = \sum_{q_i \in Q_j} \frac{r_i}{|Q_j|} \tag{5.10}$$

Given the bucket structures H_T and H_R of a test and a reference border, respectively, we are now interested in finding the best overlap between the two borders. This is performed by finding the shift $\bar{t}$ that minimize the sum of the element-wise absolute difference ($\psi(t)$) between H_R and the periodic repetition of H_T. More formally:

$$\bar{t} = \arg\min_t \psi(t) \tag{5.11}$$

with:

$$\psi(t) = \sum_{i=0}^{m-1} |H_T(i-t) - H_R(i)| \tag{5.12}$$

assuming

$$H_T(i+m) = H_T(i) \quad \forall\, i \tag{5.13}$$

The optimal values for translation $\mathbf{T} = (T_x, T_y)$ and rotation θ that lead to the best overlap of B_T and B_R according to our model are then:

$$\mathbf{T} = \mathbf{C}_R - \mathbf{C}_T \tag{5.14}$$

$$\theta = \frac{2\pi\bar{t}}{m} \tag{5.15}$$

The border matching primitives additionally returns a uniqueness U and a normalized distance DN parameters that can be used to compute a confidence threshold. The uniqueness is defined as the ratio of the second smallest minimum cost $\psi(\tilde{t})$ and the minimum cost $\psi(\bar{t})$. Formally:

$$U = \frac{\psi(\tilde{t})}{\psi(\bar{t})} \tag{5.16}$$

with:

$$\tilde{t} = \arg\min_t \psi(t) \quad t \neq \bar{t} \tag{5.17}$$

The normalized distance DN is defined as:

$$U = \frac{\psi(\bar{t})}{\sum_{i=0}^{m-1} H_T(i)} \tag{5.18}$$

5.3.2 Multi-trial approach

Trial 1 (I'm feeling lucky) : Single Border Comparison

In this first trial only a single border is computed for the test and reference images. To compute the threshold to be used for the border extraction, the

cluster centers M_ℓ and M_s are computed as shown in Subsection 3.2.4. The threshold is selected as the histogram bin with the minimum value between M_ℓ and M_s. This histogram threshold is typically different from the threshold F used for the lesion segmentation, but we verified that it is more reliable for registration purposes.

The border is then smoothed using the primitive shown in Subsection(5.3.1 with $\sigma = 20$ and finally the border matching primitive is run. The resulting match needs a sufficient uniqueness value ($U > \alpha$) and a normalized distance score ($DN < \beta$). In our test we used $\alpha = 1.2, \beta = 500$. If one of these conditions is not met then trial 2 is run, performing a more accurate comparison.

Trial 2: Multi border comparison

The second trial performs a more accurate comparison between the test and reference images. It is necessary when the comparison using the deepest minimum as the threshold (Trial 1) failed. Frequently this is due to the presence of a few local minima in the histograms with similar values, so a small difference between the two images can switch the position of the absolute minimum. In light of this, in the second trial 3 different minima are selected in the histogram and used as thresholds of the border extraction primitive. From the test image T the set of borders $\beta_T = \{B_T^1, B_T^2, B_T^3\}$ is computed; similarly, from the reference R the borders $\beta_R = \{B_R^1, B_R^2, B_R^3\}$ are obtained.

The border matching is performed between each pair in the Cartesian product $\beta_T \times \beta_R$. For each pair the confidence test is performed as in Trial 1: ($U > \alpha$) and ($DN < \beta$). If at least one pair meets these constraints a match is found and the Trial 3 test is not necessary. If more than one pair fulfills the constraints, the pair with the highest uniqueness U is used to compute the translation $\mathbf{T} = (T_x, T_y)$ and rotation θ parameters.

Trial 3: Bruteforce border comparison

This last trial uses a vast set of borders for the comparison, without any particular heuristic for the threshold selection. This step is particularly useful when the histograms show a lack of significative deep minima or the shapes of the histograms are noticeably different.

To define the two sets of thresholds for the test and the reference images,

we use the two cluster centers M_ℓ and M_s as defined in Trial 0. For the test image, given the histogram bins for the cluster centers M_ℓ^T and M_s^T, we use as the thresholds the bins between M_ℓ^T and M_s^T with a step of 2. For the reference image, with cluster centers M_ℓ^R and M_s^R, we use the same approach but with a step of 4. This leads to a cardinality for the two sets of borders β_T, β_R equals to:

$$|\beta_T| = \left\lfloor \frac{M_s^T - M_\ell^T}{2} \right\rfloor + 1 \tag{5.19}$$

$$|\beta_R| = \left\lfloor \frac{M_s^R - M_\ell^R}{4} \right\rfloor + 1 \tag{5.20}$$

As in the previous trial, the border matching is performed between each pair in the Cartesian product $\beta_T \times \beta_R$. For each pair the confidence parameters i.e. the uniqueness U and the normalized distance DN are computed. All the pairs with $(U > \alpha)$ and $(DN < \beta)$ are put into a set γ which represents all the matches with a sufficient confidence. The translation parameter $\mathbf{T} = (T_x, T_y)$ is computed as the average translation of the pair that belongs to γ. The rotation θ is instead computed as the median rotation of those pairs. If $\gamma = \emptyset$ the algorithm returns a *mismatch condition*, which means that no match with an acceptable confidence level has been found.

5.4 Experimental evaluation

In this section the algorithm performance will be evaluated in terms of accuracy and computational time.

Subsection 5.4.1 briefly describes the experimental setup. In Subsection 5.4.2 the accuracy of the registration will be compared to the registration performed by a human operator. It will be evaluated for the 3 trials, analyzing the error in the estimation of the alignment parameters and the average number of pictures that do not meet the required confidence threshold.

Subsection 5.4.3 analyzes the computational time for the 3 trials independently and for the overall algorithm. The performance is measured on a PC and a mobile device and will be compared with the constraints required by Mole Mapper.

5.4.1 Experimental Setup

The dataset is composed of 140 images of 35 different melanocytic lesions with resolution 3264×2448. Each lesion has been acquired 4 times on 4 different days. The tests are performed using the image taken on the first day as a reference and registering to it the images taken on the other days.

We use the registration performed by a human operator as the ground truth. This is obtained using a multi-touch screen showing the reference and test images in the same area. The two images are plotted alternately with a frequency of 10 Hz and the human operator can adjust the translation and rotation of the test image using two-finger gestures. Additional controls are provided for a more precise adjustment of rotation and translation independently.

5.4.2 Accuracy

	e_T (avg)	e_T (std)	e_θ (avg)	e_θ (std)
Trial 1	1.95%	2.54%	1.48°	1.24°
Trial 2	1.95%	1.31%	1.41°	1.58°
Trial 3	3.22%	1.95%	1.86°	1.46°
Total	2.08%	2.36%	1.51°	1.32°
Human operator	5.07%	3.94%	1.79°	1.58°

Table 5.1: Errors e_T,e_θ on the estimation of translation and rotation parameters (average and standard deviation) by the three trials independently, the overall algorithm and a human operator.

	Execution required	Meets confidence threshold
Trial 1	100%	74.1%
Trial 2	25.9%	60.6%
Trial 3	10.2%	100%

Table 5.2: For each trial, the fraction of elements in the dataset that require its execution and the fraction that meets the confidence threshold.

We estimated the errors e_T,e_θ of the translation and rotation parameters, respectively. The estimation of the error in translation parameters is dependent on the fixed point used for the rotation. In addition, we wanted to perform a

normalization of the error, depending on the lesion's size. For these reasons we decided to compute the translation error e_T as the ratio between the translation error on the lesion's center and the average radius of the lesion itself. More formally, we define the error as:

$$e_T = \frac{d(C'_T, C'_H)}{r}$$

where

- $d(\cdot)$ is the Euclidean distance
- C'_T is the position of the lesion's center of mass after applying the transformation estimated by the tested algorithm
- C'_H is the position of the lesion's center of mass after applying the ground truth transformation
- r is the average radius of the lesion's mask measured on the segmentation performed by MEDS algorithm (Chapter 3)

The error on the rotation e_r is computed as the difference in arc degree between the rotation estimated by the algorithm under test and the rotation imposed by the human operator while producing the ground truth.

All the issues discussed in Subsection 5.2.1 and the skin deformation in particular make the registration process a non-trivial task for a human operator also, which leads to a divergence in the registration performed by two different human operators. For this reason we compared the performance of our algorithm with the registration performed by a second human operator, using the first human operator as the ground truth.

For each trial independently we measured the error when the confidence threshold was met (see Fig. 5.1) and we observed that it was comparable with the error performed by the human operator. Similar consideration applies to the execution of the entire registration algorithm.

We finally analyzed the fraction of our dataset that meets the confidence constraints for the different trials (see Fig. 5.2) and how frequently the execution of each trial is required. On our dataset, Trial 2 is executed only 25% of the time and Trial 3 less than 11%. Additionally, every lesion on Trial 3 meets the confidence threshold so no *mismatch* condition occurs.

5.4.3 Computational resources

	PC X86		Android ARM	
	Absolute	Amortized	Absolute	Amortized
Trial 1	23	23	152	152
Trial 2	77	20	498	129
Trial 3	834	84	5638	568
Total	934	128	6288	849

Table 5.3: Average execution time (ms) on an Intel i7-4700MQ and a Qualcomm Snapdragon 800 of the three trials independently and the overall algorithm.

We measured the execution time of our registration algorithm on an X86 PC and an Android ARM device. The PC was equipped with an Intel i7-4700MQ CPU, 16GB DDR3 1600MHz and running Windows 8.1 64-bit. The ARM device was a LG Nexus 5 equipped with a Qualcomm Snapdragon 800 with 2 GB of LPDDR3-1600 RAM, running Android 4.4.2. We measured the execution time for each of the trials and the total average execution time. Additionally, we measured the contribution of each step amortized on the whole dataset, allowing a better understanding of where the time is actually spent in an average scenario.

Table 5.3 shows the measured values. The first step is always performed and requires 23ms. The second step is performed for 26% of lesions, requiring 77ms, which leads to 20ms spent on average on each lesion. The last trial requires a significant amount of time (834 ms) but is required only for $\sim 10\%$ of the lesions, leading to an amortized contribution of 84 ms.

A worst case execution time of over 6s on a mobile device might seem incompatible with the requirement of having a responsive interaction with the user, but this is not the case for the workflow actually implemented on the Mole Mapper. In fact, the automatic registration phase is performed right after the dermatoscopic image acquisition. As defined in the standard Mole Mapper workflow (see Chapter 2 Figure 2.7) the evaluation of dermatoscopic images is performed when the acquisition of all the lesions is finished. This behavior makes the proposed algorithm a viable solution since there is no constraint on the required time for the execution on a single image but only on the entire set of the acquired images.

5.5 Conclusions

We developed an algorithm for the registration of dermatoscopic images composed of three consecutive trials with increasing effectiveness and execution time.

Each trial attempts to find the best rotation-translation for aligning the test image to the reference image, comparing one or more borders from both images. When a trial finds a result with a sufficient confidence, the program stops. The first trial uses only a single border from each image; the second trials take into account three borders each, and the last trials tens of borders.

Since the problem is poorly addressed in the literature and there is no publicly available software to perform this task, we measured the absolute performance of our proposal without a comparison with other automatic registration software.

From the performed tests, our approach seems to be as accurate as the registration performed by a human operator in estimating both the parameters: translation and rotation.

The proposed method is also sufficiently fast in an average case, requiring only a fraction of a second on a Desktop PC. Although the worst case may require several seconds, the standard workflow implemented on Mole Mapper allows us to focus only on the average performance, since all of the registrations are performed in the background while the dermatologist is taking the pictures.

Chapter 6

Conclusions

We developed Mole Mapper in strict collaboration with the Dermatological Clinic of Padova to increase the levels of accuracy, efficiency and confidence for melanoma diagnosis. It combines a carefully designed interface with advanced image processing algorithms.

Among the several computer vision algorithms developed for Mole Mapper, three major examples, working on dermatoscopic images, have been deeply described and analyzed because of their scientific relevance and the centrality of their roles in Mole Mapper. These algorithms are: Melanocytic lesion segmentation; Virtual Shave; and Dermatoscopic images registration.

The Melanocytic lesion segmentation is based on a threshold scheme which mimics the cognitive process of human dermatologists. It outperforms the accuracy of state-of-the-art techniques; furthermore, it has been proven to be more accurate than dermatologists with little - or even moderate - experience in the field, and only slightly worse than the most experienced dermatologists. It is also extremely fast, requiring only a fraction of seconds on handheld devices, demonstrating an improvement of an order of magnitude or more over state-of-the-art methods. We have additionally proved that our segmentation is robust with regard to the most common photographic defects, light variation, and it is highly reproducible in the presence of skin deformation.

Virtual Shave performs hair removal on dermatoscopic images. The detection

phase is performed by merging the information coming from multiple DoG filters followed by the analysis of the characteristics of the connected components. The repair phase is optional and used only for visualization purposes. Depending on the requirements, bilinear interpolation or a Telea fast-marching inpainting can be used. We compared our proposed technique with the well-known and freely available DullRazor software. Our method outperformed DullRazor, achieving a massive accuracy improvement with almost one-third of the error rate. The technique is also very fast, performing approximately 30 times faster than the implementation of DullRazor.

Dermatoscopic images registration addresses a problem which has been poorly addressed in the literature. We performed an analysis and characterization of the problem and proposed a 3-trials algorithm that relies on the robustness of MEDS segmentation. The different trials have increasing effectiveness and computational cost and the algorithm stops as soon as a sufficient confidence level is reached. Our technique appears to be as accurate as a human operator in evaluating translation and rotation parameters. The algorithm is fast enough for our purposes since it requires only a fraction of a second for an average case. The specific workflow implemented on Mole Mapper hides the higher time required by the most difficult cases, avoiding any slowdown during the visit process.

Bibliography

[1] R. Morris-Jones, *ABC of Dermatology, 6th Edition*. Wiley, 2014.

[2] A. C. Society., "Sun protection. cancer trends progress report - 2011/2012 update." `http://progressreport.cancer.gov/doc_detail.asp?pid=1&did=2007&chid=71&coid=711&mid`. [Online; Accessed December, 2014].

[3] R. S. Stern, "Prevalence of a history of skin cancer in 2007: results of an incidence-based model," *Arch Dermatol*, vol. 146, no. 3, pp. 279–282, 2010.

[4] A. C. Society., "Cancer facts and figures 2014.." `http://www.cancer.org/acs/groups/content/@research/documents/webcontent/acspc-042151.pdf`. [Online; Accessed December, 2014].

[5] N. Howlader, A. Noone, M. Krapcho, *et al.*, "SEER Cancer Statistics Review, 1975-2011 based on november 2013 seer data submission. Bethesa, MD: National Cancer Institute," 2014. [Online; Accessed December, 2014].

[6] A. V. Giblin and J. M. Thomas, "Incidence, mortality and survival in cutaneous melanoma," *J Plast Reconstr Aesthet Surg*, vol. 60, no. 1, pp. 32–40, 2007.

[7] W. Liu, J. Dowling, W. Murray, *et al.*, "Rate of growth in melanomas: Characteristics and associations of rapidly growing melanomas," *Archives of Dermatology*, vol. 142, no. 12, pp. 1551–1558, 2006.

[8] E. W. Breitbart, A. Waldmann, S. Nolte, M. Capellaro, R. Greinert, B. Volkmer, and A. Katalinic, "Systematic skin cancer screening in Northern Germany," *J. Am. Acad. Dermatol.*, vol. 66, no. 2, pp. 201–211, 2012.

[9] L. A. Pollack *et al.*, "Melanoma survival in the United States, 1992 to 2005," *J. Am. Acad. Dermatol.*, vol. 65, no. 5 Suppl 1, pp. 78–86, 2011.

[10] M. L. Bafounta, A. Beauchet, P. Aegerter, and P. Saiag, "Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests," *Arch Dermatol*, vol. 137, no. 10, pp. 1343–1350, 2001.

[11] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *Lancet Oncol.*, vol. 3, no. 3, pp. 159–165, 2002.

[12] M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *Br. J. Dermatol.*, vol. 159, no. 3, pp. 669–676, 2008.

[13] J. Malvehy, S. Puig, G. Argenziano, A. A. Marghoob, and H. P. Soyer, "Dermoscopy report: proposal for standardization. Results of a consensus meeting of the International Dermoscopy Society," *J. Am. Acad. Dermatol.*, vol. 57, no. 1, pp. 84–95, 2007.

[14] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types I through VI," *Arch Dermatol*, vol. 124, no. 6, pp. 869–871, 1988.

[15] D. Schadendorf, C. Kochs, and E. Livingstone, *Handbook of Cutaneous Melanoma, A Guide to Diagnosis and Treatment*. Springer, 2013.

[16] N. R. Abbasi, "Early Diagnosis of Cutaneous Melanoma: Revisiting the ABCD Criteria," *Jama-journal of The American Medical Association*, vol. 292, pp. 2771–2776, 2004.

[17] C. Piccolo, "A low-cost, efficient and accurate human body scanner," 2012.

[18] I. Bogdan, J. Smolle, H. Kerl, G. Burg, and R. Boni, "Melanoma ex naevo: a study of the associated naevus," *Melanoma Res.*, vol. 13, no. 2, pp. 213–217, 2003.

[19] M. Bonazza, "A workflow for skin cancer evaluation," 2015.

[20] F.Bogo, F.Peruch, A. Fortina, and E. Peserico, "Variability in human and automated segmentation of dermoscopy images of melanocytic skin lesions," *Dermoscopy Image Analysis, M.E. Celebi, T. Mendonca, J.S. Marques eds, CRC Press/Taylor Francis*, To appear.

[21] F. Nachbar, W. Stolz, T. Merkle, A. Cognetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy," *JAAD*, vol. 30, no. 4, pp. 551–559, 1994.

[22] G. Joel *et al.*, "Validation of segmentation techniques for digital dermoscopy," *Skin Res. & Tech.*, vol. 8, no. 4, pp. 240–249, 2002.

[23] A. Silletti, E. Peserico, A. Mantovan, E. Zattra, A. Peserico, and A. Belloni Fortina, "Variability in human and automatic segmentation of melanocytic lesions," in *Proc. of IEEE EMBC*, pp. 5789–5792, 2009.

[24] A. Belloni Fortina, E. Peserico, A. Silletti, and E. Zattra, "Where's the naevus? Inter-operator variability in the localization of melanocytic lesion border," *Skin Res. & Tech.*, vol. 18, no. 3, pp. 311–315, 2011.

[25] F. Peruch, F. Bogo, M. Bonazza, V. Cappelleri, and E. Peserico, "Simpler, faster, more accurate melanocytic lesion segmentation through meds," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 2, pp. 557–565, 2014.

[26] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, vol. 56, no. 2, pp. 69–90, 2012.

[27] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Res. & Tech.*, vol. 11, no. 1, pp. 17–26, 2005.

[28] H. Zhou, G. Schaefer, M. E. Celebi, F. Lin, and T. Liu, "Gradient vector flow with mean shift for skin lesion segmentation," *Comput. Med. Imaging Graph.*, vol. 35, no. 2, pp. 121–127, 2011.

[29] M. E. Celebi, H. Iyatomi, G. Schaefer, and W. V. Stoecker, "Lesion border detection in dermoscopy images," *Comput. Med. Imaging Graph.*, vol. 33, no. 2, pp. 148–153, 2009.

[30] M. Silveira, J. Nascimento, J. Marques, A. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, 2009.

[31] M. E. Celebi, Y. Aslandogan, and W. V. Stoecker, "Unsupervised border detection in dermoscopy images," *Skin Res. & Tech.*, vol. 13, no. 4, pp. 454–462, 2007.

[32] J. Gao, J. Zhang, M. G. Fleming, I. Pollak, and A. B. Cognetta, "Segmentation of dermatoscopic images by stabilized inverse diffusion equations," in *Proc. of ICIP*, pp. 823–827, 1998.

[33] M. E. Celebi *et al.*, "Border detection in dermoscopy images using Statistical Region Merging," *Skin Res. & Tech.*, vol. 14, no. 3, pp. 347–353, 2008.

[34] H. Wang *et al.*, "Watershed segmentation of dermoscopy images using a Watershed technique," *Skin Res. & Tech.*, vol. 16, no. 3, pp. 378–384, 2010.

[35] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[36] M. E. Celebi, Q. Wen, S. Hwang, H. Iyatomi, and G. Schaefer, "Lesion border detection in dermoscopy images using ensembles of thresholding methods," *Skin Res. & Tech.*, vol. 19, no. 1, pp. e252–e258, 2013.

[37] F. Peruch, F. Bogo, M. Bonazza, M. Bressan, V. Cappelleri, and E. Peserico, "Simple, fast, accurate melanocytic lesion segmentation in 1D colour space," in *Proc. of VISAPP*, pp. 191–200, 2013.

[38] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved Internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Comput. Med. Imaging Graph.*, vol. 32, no. 7, pp. 566–579, 2008.

[39] D. Delgado, C. Butakoff, B. K. Ersboll, and W. V. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Trans. on Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, 2008.

[40] R. Melli, C. Grana, and R. Cucchiara, "Comparison of color clustering algorithms for segmentation of dermatological images," in *Proc. of SPIE*, 2006.

[41] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering comparison," *IEEE Trans. on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[42] R. Cucchiara, C. Grana, S. Seidenari, and G. Pellacani, "Exploiting color and topological features for region segmentation with recursive Fuzzy c-means," *Machine Graph. and Vision*, vol. 11, no. 2/3, pp. 169–182, 2002.

[43] H. Zhou, G. Schaefer, A. Sadka, and M. E. Celebi, "Anisotropic mean shift based Fuzzy c-means segmentation of dermoscopy images," *Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 26–34, 2009.

[44] H. Zhou, M. Chen, R. Gass, J. M. Rehg, L. Ferris, J. Ho, and L. Drogowski, "Feature preserving artifact removal from dermoscopy images," in *Proc. of SPIE*, 2008.

[45] P. Wighton, T. K. Lee, and M. S. Atkins, "Dermascopic hair disocclusion using inpainting," in *Proc. SPIE 6914, Medical Imaging 2008: Image Processing*, 2008.

[46] M. Fiorese, E. Peserico, and A. Silletti, "VirtualShave: Automated hair removal from digital dermatoscopic images," in *Proc. of IEEE EMBC*, pp. 5145–5148, 2011.

[47] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Comp. Stat.*, vol. 2, no. 4, pp. 433–459, 2010.

[48] S. Nakariyakul, "Fast spatial averaging: an efficient algorithm for 2D mean filtering," *The Journal of Supercomputing*, vol. 65, no. 1, pp. 262–273, 2013.

[49] K. Suzuki, I. Horiba, and N. Sugie, "Linear-time connected-component labeling based on sequential local operations," *Comput. Vis. Image Underst.*, vol. 89, no. 1, 2003.

[50] F. Chang, C.-J. Chen, and C.-J. Lu, "A linear-time component-labeling algorithm using contour tracing technique," *Comput. Vis. Image Underst.*, vol. 93, no. 2, pp. 206–220, 2004.

[51] J.-M. Park, C. G. Looney, and H.-C. Chen, "Fast connected component labeling algorithm using a divide and conquer technique," in *Proc. of ISCA*, pp. 373–376, 2000.

[52] J. Martin-Herrero, "Hybrid object labelling in digital images," *Machine Vision and Applications*, vol. 18, no. 1, pp. 1–15, 2007.

[53] FotoFinder Systems Inc., "Fotofinder dermoscope." http://www.fotofinder.de/en.html.

[54] M. Wei-Ying and B. S. Manjunath, "EdgeFlow: a technique for boundary detection and image segmentation," *IEEE Trans. on Image Processing*, vol. 9, no. 8, pp. 1375–1388, 2000.

[55] 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, no. 1, pp. 104–111, 1996.

[56] T. Lee *et al.*, "Dullrazor: A software approach to hair removal from images.," *Computers in Biology and Medicine*, 1997.

[57] X. Feng-Ying, S.-Y. Qin, Z.-G. Jiang, and R.-S. Meng, "Pde-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma.," *Comp. Med. Imag. and Graph.*, vol. 33, no. 4, pp. 275–282, 2009.

[58] P. Schmid-Saugeona, J. Guillodb, and J. P. Thirana, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Comput Med Imaging Graph*, vol. 27, no. 1, pp. 65–78, 2003.

[59] S. A. Fiorese M., Peserico E., "Virtualshave: automated hair removal from digital dermatoscopic images," *Conf Proc IEEE Eng Med Biol Soc*, vol. 2011, pp. 5145–8, 2011.

[60] A. Afonso and M. Silveira, "Hair detection in dermoscopic images using percolation," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 4378–4381, 2012.

[61] K. Kiani and A. R. Sharafat, "E-shaver: An improved dullrazor® for digitally removing dark and light-colored hairs in dermoscopic images," *Comput. Biol. Med.*, vol. 41, no. 3, pp. 139–145, 2011.

[62] N. Nguyen, T. Lee, and M. Atkins, "Segmentation of light and dark hair in dermoscopic images: a hybrid approach using a universal kernel," in *Proc. SPIE 7623, Medical Imaging 2010: Image Processing*, 2010.

[63] Toossi *et al.*, "An effective hair removal algorithm for dermoscopy images. skin research and technology," *Skin Research and Technology*, pp. 230–235, 2013.

[64] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, 1986.

[65] M. G. Fleming, C. Steger, J. Zhang, J. Gao, A. B. Cognetta, I. Pollak, and C. R. Dyer, "Techniques for a structural analysis of dermatoscopic imagery," *Comput Med Imaging Graph*, vol. 22, no. 5, pp. 375–389, 1998.

[66] H. Zhou, M. Chen, R. Gass, J. M. Rehg, L. Ferris, J. Ho, and L. Drogowski, "Feature preserving artifact removal from dermoscopy images," in *Proc. of SPIE Medical Imaging: Image Processing*, (San Diego, USA), 2008. Accepted for publication.

[67] C. Steger, "An unbiased detector of curvilinear structures," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 2, pp. 113–125, 1998.

[68] Q. Abbas, I. Fondon, and M. Rashid, "Unsupervised skin lesions border detection via two-dimensional image analysis," *Computer Methods and Programs in Biomedicine*, pp. 1–1, 2011.

[69] Q. Abbas, M. E. Celebi, and I. F. García, "Hair removal methods: A comparative study for dermoscopy images," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 395–404, 2011.

[70] Q. Abbas, I. F. Garcia, M. Emre Celebi, and W. Ahmad, "A feature-preserving hair removal algorithm for dermoscopy images," *Skin Res Technol*, vol. 19, no. 1, pp. 27–36, 2013.

[71] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, 1990.

[72] C. A. Z. Barcelos and V. B. Pires, "An automatic based nonlinear diffusion equations scheme for skin lesion segmentation.," *Applied Mathematics and Computation*, vol. 215, no. 1, pp. 251–261, 2009.

[73] D. H. Chung and G. Sapiro, "Segmenting skin lesions with partial differential equations based image processing algorithms.," in *ICIP*, pp. 404–407, 2000.

[74] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *Trans. Img. Proc.*, vol. 13, no. 9, pp. 1200–1212, 2004.

[75] F. Bornemann and T. März, "Fast image inpainting based on coherence transport," *J. Math. Imaging Vis.*, vol. 28, no. 3, pp. 259–278, 2007.

[76] F. I. Abbas Q., Emre Celebi M., "Computer-aided pattern classification system for dermoscopy images," *Skin Research and Technology*, pp. 278–289, 2012.

[77] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," *Commun. ACM*, vol. 32, no. 3, pp. 359–373, 1989.

[78] A. Telea, "An image inpainting technique based on the fast marching method.," *J. Graphics, GPU and Game Tools*, vol. 9, no. 1, pp. 23–34, 2004.

[79] S. Pavlopoulos, "New hybrid stochastic-deterministic technique for fast registration of dermatological images," *Medical and Biological Engineering and Computing*, vol. 42, no. 6, pp. 777–786, 2004.

[80] G. Maletti and B. Ersbøll, "A combined alignment and registration scheme of psoriasis lesion images," in *IMM-Technical Report-2003-9*, (Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby), p. 42, 2003.

[81] D. Delgado, B. Ersbøll, and J. M. Carstensen, "S.h.a.r.p: A smart hierarchical algorithm to register psoriasis," in *International workshop on Systems, Signals and Image Processing*, pp. 43–46, 2004.

[82] S. Zambanini, G. Langs, R. Sablatnig, and H. Maier, "Automatic robust registration of cutaneous hemangiomas for follow-up examinations," in *Proc. of 31st AAPR/OAGM Workshop*, 2007.

[83] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.

[84] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[85] C.-N. Anagnostopoulos, D. D. Vergados, and P. Mintzias, "Image registration of follow-up examinations in digital dermoscopy.," in *BIBE*, pp. 1–4, IEEE, 2013.

[86] F. Mokhtarian and A. K. Mackworth, "Scale-based description and recognition of planar curves and two-dimensional shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 1, pp. 34–43, 1986.

List of Tables

List of Figures