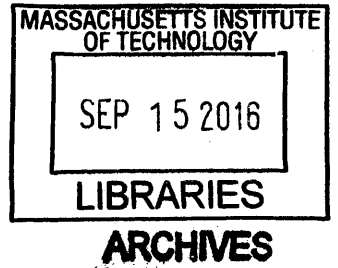# Computational Tools for Enabling Longitudinal Skin Image Analysis

by

Kang Qi Ian Lee

BEng, Electrical and Electronic Engineering
Imperial College London (2014)

Submitted to the Center for Computational Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Computation for Design and Optimization

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016

Author .... **Signature redacted** ....
Center for Computational Engineering
August 5, 2016

Certified by **Signature redacted** ....
Brian W. Anthony
Principal Research Scientist, Department of Mechanical Engineering
Thesis Supervisor

Accepted by ..... **Signature redacted** ......
Youssef M. Marzouk
Associate Professor, Department of Aeronautics and Astronautics
Co-Director, Computation for Design and Optimization

MITLibraries

DISCLAIMER NOTICE

Computational Tools for Enabling Longitudinal Skin Image Analysis

by

Kang Qi Ian Lee

Submitted to the Center for Computational Engineering
on August 5, 2016, in partial fulfillment of the
requirements for the degree of
Master of Science in Computation for Design and Optimization

Abstract

We present a set of computational tools that enable quantitative analysis of longitudinally acquired skin images: the assessment and characterization of the evolution of skin features over time. A framework for time-lapsed skin imaging is proposed. A nonrigid registration algorithm based on multiple plane detection for landmark identification accurately aligns pairs of longitudinal skin images. If dense and thick hairs are present, then nonrigid registration is used to reconstruct the skin texture of occluded regions by recording multiple images from the same area. Realistic reconstruction of occluded skin texture is aided by an automatic hair segmentation algorithm and guided painting method based on image blending. We demonstrate that constituent algorithms in this framework are accurate and robust in a multitude of scenarios. In addition, a methodology for rigorous longitudinal analysis of skin microrelief structure is introduced. Following rigid registration, a microrelief junction point matching algorithm based on point pattern matching is shown to accurately match two sets of junction points. Immediate applications for these computational tools are change detection for pigmented skin lesions and deformation field computation of the skin surface under stress using only visual features of the skin. Prospective applications include new insights in skin physiology and diseases from the capability to precisely track movements of the microrelief structure over time and localization of skin images on the body.

Thesis Supervisor: Brian W. Anthony
Title: Principal Research Scientist, Department of Mechanical Engineering

Acknowledgments

After 2 years at MIT, I am quite reluctant to leave the community and the memorable experiences behind. The journey has been nothing short of amazing. And I have many people to thank for.

The first person I must thank is my advisor, Dr. Brian Anthony. I would always come out of meetings with him with more ideas. So he has been a great help towards helping me find my research direction in skin research. There was a lot of uncertainty in exploring this field of study and with his guidance, we found interesting areas to work on. From him I learned to: try out things, and avoid thinking too much; always be ready to present; appreciate his sarcasm when he jokes with fellow lab mates. So working with him has been immensely pleasurable.

Next, I would like to thank my colleague Dr. Xian Du. He knows everything; name any field of engineering and he has done something on it. So I am thankful to have him talk about algorithms, paper writing and his current projects at the interface of so many fields. Most importantly, I am grateful for our collaboration in skin research that was made into part of this thesis. Also, as a friend and colleague in Team Skin, I thank Ina Kundu for her hard work in designing the imaging system used in this thesis and for meticulously acquiring the skin images I needed.

Over the past year, the lab doubled in size and I really enjoyed all the friendships and fun activities we had. A special mention goes to Rishon and Alex Benjamin, who let me stay in their Tang apartment while I finish writing this thesis. It is great to have fellow CDO coursemates in the lab. We are a rare breed.

Being part of CDO brought me into unfamiliar territory when its came to classes. As such, I am sincerely grateful for my friends from our program who supported me through the toughest psets and projects in the core CDO classes, hung out for countless hours in the old CCE office and went for hikes in and around Boston. Special thanks go to Ming Qing Foo, Lutao Xie, Park Sinchaisri, Larson Hogstrom and Ken Wang.

More thanks go to Lezhi Li and Wendy Wei, who were the most wonderful team-

mates I ever worked with in class. I am truly glad that our 6.869 project kickstarted Lezhi's journey into discovering the visual identity of cities.

I thank the Singaporeans in MIT for all the fun and good times, meals together, board games and travels. Special mention goes to Goh Siongthye, Kong Jian Feng, Dax Koh, Gladia Hotan, Chen Hanrong, Low Guang Hao, Alvin Tan, Tan Zheng Jie, Liu Yun and Yeong Li Qian.

My heartfelt thanks go to my small group, who has been a great source of love and support throughout my time here: Ryan Chang, Alice Kao, Wu Mengfei, David Kwabi, Lyndon Zhang, Tim Yee and Kim MinJung.

My sincere gratitude go to my good friends in Graduate Christian Fellowship: Yong Sze Zheng, Nigel Chou, Grace Goon, Sam Elder and Kevin Meng.

Finally, I would like to thank my parents and brother for their love and support, whom I have spent 5 years away from due to my university studies.

Contents

List of Figures

11

13

14

List of Tables

Chapter 1

Introduction

A visual inspection of human skin reveals many salient visual features. Variations in color pigmentation, hair shafts and hair follicle openings are immediately noticeable. There are numerous intersecting fine ridges appear like irregular geometric patterns; these are the microrelief or glyphic patterns of the skin. Wrinkles appear alongside the microrelief patterns or deepen their appearance. Many of these skin features correspond to visual features that are quite uniformly distributed in visible light optical images of the skin.

The field of computational skin analysis is traditionally concerned with problems such as computer-aided diagnosis of skin cancer from skin images [47] and extracting and matching dermatoglyphic patterns on fingerprints for biometric identification [105]. To the best of our knowledge, no work has quantitatively assessed and characterized the evolution of skin features over time. A nonrigid skin image registration would be required to address this gap in research. In this thesis, we describe an accurate and robust skin image registration algorithm that first identifies and match visual features of the skin between image pairs. Then, a local motion model uses these corresponding landmarks to register the skin images. An accurate non-rigid registration method certainly has some novel applications in computational skin analysis. For example, change detection of a pigmented skin lesion (PSL) [47] could be accurately determined using a freehand imaging device. Also, the capability to track the movements of the microrelief structure over time can potentially lead to new insights

17

in skin physiology, pathological analysis and disease diagnosis. Another possible application for the algorithm is in finding the dense deformation field of the skin surface under stress without using physical markers. Currently, this measurement is obtained using markers on the skin [64]. Moreover, there are potential uses for registration in localizing of skin images on the body; to elaborate, the skin surface could be utilized as a map for determining 3D ultrasound probe pose [92].

Accurate segmentation of pathological skin features is a major research area in computational skin analysis. For instance, segmentation of a PSL allow algorithms to ascertain whether melanoma is present using the ABCDE criteria [47]: The shape of the segmentation is used to assess Asymmetry, Border, Color, Diameter and Evolving characteristics of a mole. But oftentimes hair occlude these skin images and negatively affect the accuracy of PSL segmentation. Consequently, digital hair removal (DHR) is used as a preprocessing step to PSL segmentation [45]. In our case, we employ DHR as a preprocessing step to general skin feature extraction, including skin microrelief structure extraction. However, existing DHR techniques in the literature are optimized for removing dark hairs that often share similar colors with PSLs in dermoscope images [53, 3, 104]. Additionally, dermoscope images are recorded by pressing a piece of glass onto the skin [65], so the shallow, net-like microrelief structures are not very visible in these images. Therefore, existing DHR methods are prone to falsely identifying microrelief structure, when visible, as hairs. In this thesis, we propose a framework for realistic reconstruction of skin texture and microrelief from multiple images of the same patch of skin. First, our automatic hair segmentation algorithm segments hairs that occlude the microrelief structure of the skin images. Second, the images are registered using our nonrigid skin registration algorithm. Third, areas of the first image with missing information due to removed hairs are filled in with what we know about the skin texture in other images by guided inpainting.

A potential use for skin registration is in the assessment of the stability and evolution of microrelief structure over time. Although some studies have attempted to extract microrelief structure in order to create measurement indices for skin aging assessment [111, 43, 15], they did not attempt to track shifts in the structure over

18

time. Skin registration enables rigorous scientific inquiry into the long-term stability of microrelief structure. In this thesis, we describe a method that initially uses feature-based rigid skin registration to identify and register a region of interest (ROI) from a pair of time-lapsed skin images. After extracting junctions points from the net-like microrelief, a microrelief junction point matching (MJPM) algorithm employs nonrigid point pattern matching (PPM) to match both point sets. This is a first step towards prospective applications such as novel methods for tracking skin diseases over time and skin-based biometric identification.

1.1   Contributions

A key contribution of this thesis is the development and evaluation of enabling technologies for quantitative time-lapsed dermatology. Nonrigid skin image registration accurately aligns pairs of skin images recorded with a freehand visible light optical skin imaging system at regular intervals of time over a time scale of months. If dense and thick hairs are present in the time-lapsed images, then nonrigid skin image registration is used to reconstruct the skin texture of occluded regions by recording multiple images from the same area. Next, hairs are accurately extracted with our automatic hair segmentation algorithm and subsequently removed with guided inpainting to produce a realistic reconstruction of skin texture. Moreover, skin registration can be applied towards tracking deformations of the skin surface using only visual features of the skin.

Another major contribution is in the introduction of novel methodology for rigorous time-lapsed analysis of microrelief structure. A preliminary step we have taken is the proof-of-concept validation of MJPM for accurate matching of two sets of rigidly registered microrelief structure.

19

1.2 Thesis Outline

The thesis is structured as follows. Chapter 2 gives the reader background knowledge about the freehand optical imaging system and skin features. This is followed by a literature review on skin registration, DHR, microrelief structure extraction and non-rigid PPM. Then, Chapter 3 describes methodology for our skin registration algorithm in two parts: A feature-based homography registration algorithm gives an initial guess, and then a (MPD) registration technique finds the deformation transformation. This is followed by an evaluation of the MPD-based nonrigid registration method. In Chapter 4, a description of MJPM consisting of two parts is given. A enhanced feature-based rigid homography registration method for the initial guess and PPM algorithm for matching are detailed. Next, the performance of both rigid registration and PPM are assessed. Chapter 5 presents the automatic hair segmentation algorithm and then evaluates its performance. Afterwards, Chapter 6 reports the methodology and proof-of-concept evaluation for guided inpainting. Finally, Chapter 7 concludes this thesis with summary of results, limitations and future work.

Chapter 2

Background

In this chapter, key features of the visible light skin optical imaging system used to capture skin images are highlighted. Second, skin features relevant to design choices in this thesis are described. Third, the registration problem is defined and prior work in skin registration is discussed. Fourth, related work on digital hair removal (DHR) in terms of image segmentation and image inpainting is examined. Finally, a review of literature related to microrelief junction point matching (MJPM) is given: microrelief extraction and nonrigid point pattern matching (PPM).

2.1   Visible Light Optical Skin Imaging System

The freehand visible light optical imaging system is designed by Kundu [48]. We summarize the key design features of the system in this section.

The imaging device consists of both custom-made and off-the-shelf parts that provide high-resolution images of the skin surface so that skin features can be resolved in detail. A LabView-controlled Basler ac2040-90uc color camera provides a resolution of 26.6 µm per pixel at a working distance of 6.8 cm. Tests conducted in [48] concluded that the camera provides sufficient resolution to resolve skin features. The focal length of the lens is 16 mm. An Edmund Optics DC LED light source with a light diffuser is used to illuminate the skin surface. The lens hood is lined with a enhanced-reflectivity metallic polymer substrate on the outside to reduce the effects of ambient lighting,

while the inside of the hood is lined with white paper to ensure uniform illumination of the skin from the light source. These components are combined into a custom-made, handheld probe illustrated in Figure 2-1.

To acquire images, the human operator places the imaging device with optical axis perpendicular to the skin surface at the region of interest (ROI). An RGB image of resolution 2040×2040 is acquired and is saved in the lossless *png* format. As the device is handheld, the operator can only approximately align the region of interest (ROI) on successive images of the same region in serial examinations.
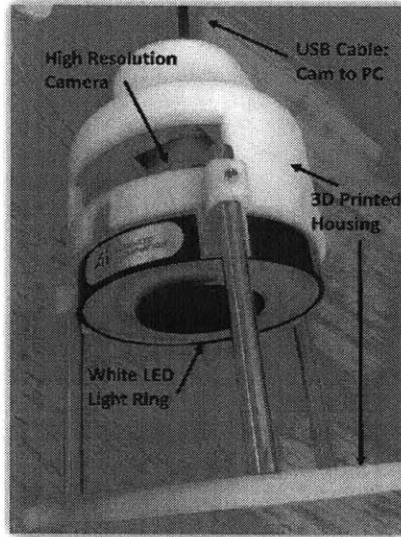
Figure 2-1: Imaging device design: a high-resolution Basler camera and LED light ring built into custom-made handheld probe.

2.2 Skin Features

The imaging system resolves skin features that are at least $26.6\,\mu m$ in length, and outputs images that are approximately $5.4\,cm \times 5.4\,cm$ in area. In this thesis, several skin features are relevant to the design choices of our algorithms: (1) skin pigmentation, (2) hair, and (3) microrelief structure. These three skin features are distributed all across the human body, but their density and appearance are variable in different parts of the body. In the rest of this section, we briefly describe the skin features and

their relation to our work.

2.2.1 Skin Pigmentation

The color of the skin arises from pigments that absorb light called chromophores. The dominant chromophore in the epidermis, the outermost layer of the skin, is melanin, which ranges in colors from black to brown (eumelanin) and red to yellow (pheomelanin). Melanin is produced by melanocytes, which are cells found in the basal layer of the epidermis. When exposed to sunlight, melanocytes produce melanin, and this biological reaction eventually causes skin to appear tanned. Another chromophore is hemoglobin, a red pigment carried in erythrocytes (red blood cells). These cells are found in the capillaries and veins of the dermis, the layer beneath the epidermis. The amount of blood flow determines the amount of reddish hue in skin [55]. The abovementioned skin layers and structures are displayed in Figure 2-2. More details on the physiology of skin pigmentation is found in [35].

Clusters of these chromophores in the skin layers give rise to salient skin features such as moles and freckles. These skin features are visible to the naked eye and occur at a scale of up to approximately 2 cm [35]. Due to variability of these features over time, they are not desirable visual features for visual landmarks in skin registration. Furthermore, some of these features are also the subject of image segmentation for computer-assisted diagnosis by change detection, and thus should not be used to register images for longitudinal analysis.

2.2.2 Hair

Hair is composed of cylindrical structures made of compacted cells that grow from hair follicles in the dermal layer, emerging from the surface of the skin. This is illustrated in 2-2. Brown and black hairs are pigmented by eumelamin, red and yellow hairs are pigmented by phomelanin and white hairs are unpigmented. In humans, the diameter of each hair shaft ranges from 15 μm to 160 μm. Human hair can be classified into 3 kinds: (1) terminal hair, (2) vellus hair and (3) intermediate
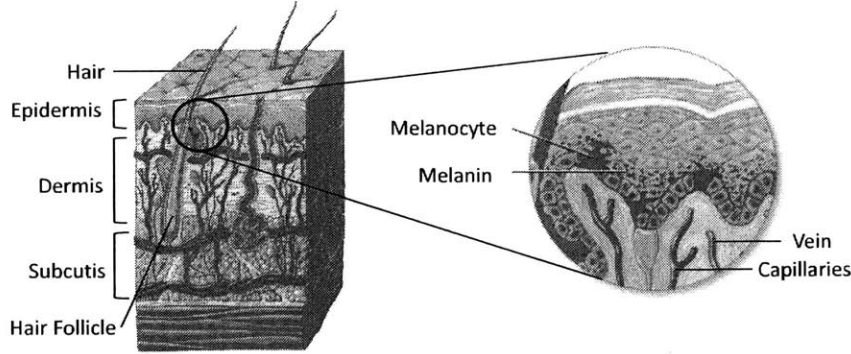
Figure 2-2: Layers of the skin and hair structure on left, and enlarged view of melanocytes, melanin and blood vessels on right. Images courtesy of A.D.A.M. [1].

hair. Terminal hairs are long, thick ($>60\,\mu$m) and pigmented, and are found in areas such as the scalp, beard and armpit. Vellus hair are short (approximately 1 mm), thin ($<30\,\mu$m) and unpigmented; they are found in areas such as the forehead and bald scalp. Intermediate hairs have intermediate lengths and diameters and are found on the legs and arms. [11, 30, 35]

When dense amounts of terminal and intermediate hairs partially obstruct the view of skin pigmentation and microrelief structure, complications occur in accurately extracting the underlying skin features. For instance, hairs cause occlusion in images of pigmented skin lesions (PSL) and without first removing the hairs, PSL segmentation algorithms often give erroneous results because hair colors may be similar to those of the lesion. [45] Consequently, to address this problem, a number of DHR algorithms were developed. A summary of these algorithms is described in Chapter 2.4. These techniques mainly exploit the property that hair is curvilinear; local operators that detect lines in images are used in order to enhance these ridge or valley structures. These algorithms perform very well on skin images with thick, dark hairs with few intersections between hairs and high contrast between hair and skin (Figure 2-3a). However, it remains challenging for algorithms to distinguish between hair and microrelief with similar contrast in intensity (Figure 2-3c), differentiate thin, faintly-colored intermediate hairs and skin (Figure 2-3d), and to discriminate dark and light hair structures (Figure 2-3e). Also, difficulty arises when hairs clump together as

24

they do not appear like curvilinear structures (Figure 2-3f). Moreover, the imaging system introduces several artifacts in hair-occluded skin images. At certain viewing angles, severe amounts of specular reflections are observed on hair (Figure 2-4a); this occassionally causes gaps in detected hair strands. In addition, hairs may be out of focus as the depth of field of the camera is small at a short working distance (Figure 2-4a). Compared to hairs in focus, out of focus hairs have larger widths and exhibit lower contrast with respect to skin, thus making them more difficult to extract. In contrast to dermoscope images [65], the imaging system does not flatten hairs onto the skin with a transparent glass plate or lens as it aims to cover a larger patch of skin. The Bayer filter RGB camera sensor also causes demosaicing artifacts [58] during image conversion (Figure 2-4b) that is made worse by high contrast between skin and hair. This makes it difficult to ascertain the true edges of the hairs.

2.2.3 Microrelief Structure

The net-like lines forming triangles and quadrilaterals on the skin surface is the microrelief structure, seen in Figure 2-5. Two types of skin lines are present: (1) primary lines, which are wide and uniformly directed, with depth $20\,\mu\text{m}$ to $100\,\mu\text{m}$; (2) secondary lines, which are thin and proceed in all other directions, with depth $5\,\mu\text{m}$ to $40\,\mu\text{m}$. [85] The appearance of the microrelief is affected by aging and environmental factors such as sun exposure and humidity. [79]

We postulate that the microrelief structure give rise to stable skin features in a time scale of months to years. Furthermore, the microrelief structure contributes most to the appearance of detailed visual features on the skin surface. Therefore, landmarks from the skin could be used to register images taken over some period of time or images taken from different viewpoints in one session.

2.3   Skin Registration

Given a target image T and source image S in the image domain Ω, and a transformation W, the aim of registration is to estimate an optimal transformation W^* that
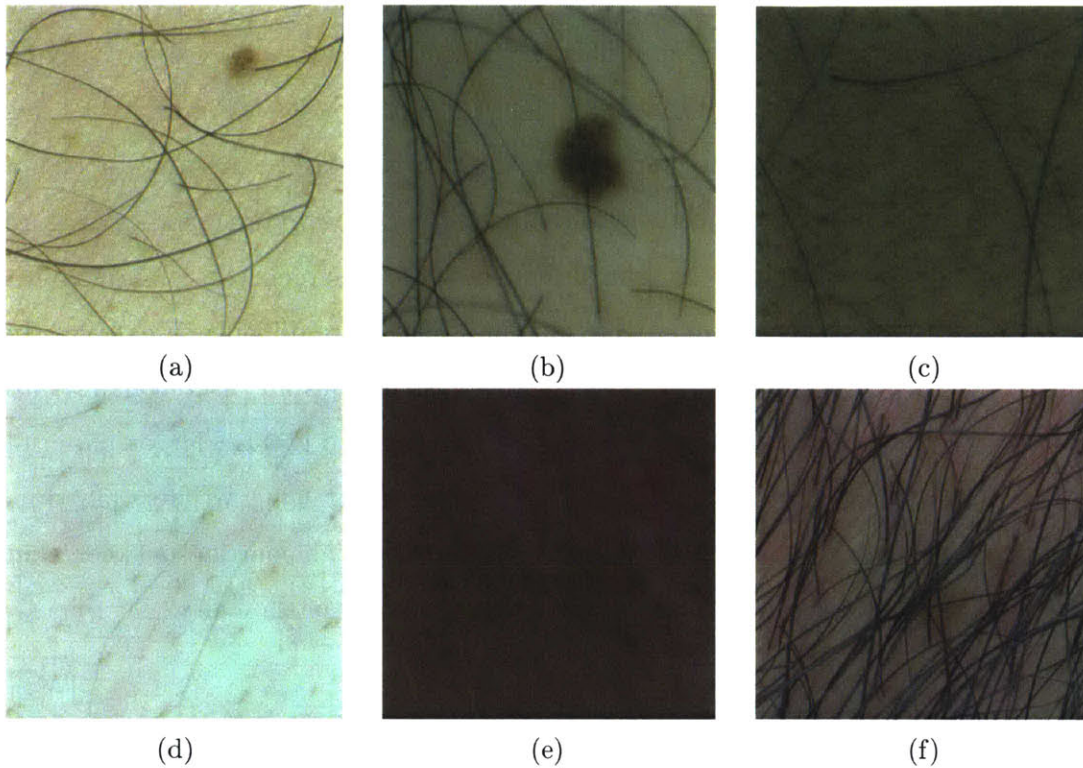
Figure 2-3: a) Thick, dark, individual hair strands with few intersections are easy to detect with existing DHR algorithms. (b-f) Hair and skin types that are difficult to detect. (b) Hairs obstructing the view of moles and microrelief structure. (c) Microrelief on skin exhibits similar contrast to that of hair on skin. (d) Thin and faint intermediate hairs. (e) Mixture of dark and light hairs. (f) Hair clumping together do not appear like curvilinear structures.
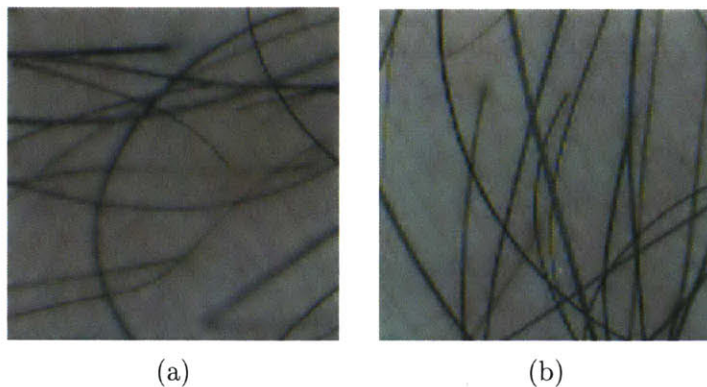
Figure 2-4: Imaging system artifacts in hair-occluded skin images. (a) Specular reflections of hair and hairs out of focus. (b) Demosaicing artifacts from Bayer filter camera sensor.
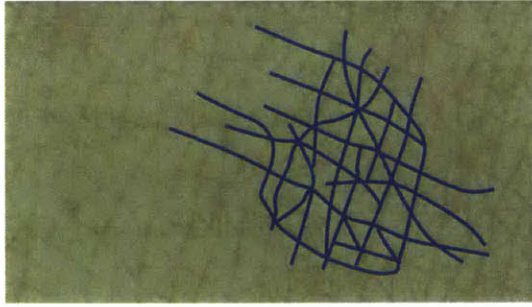
Figure 2-5: Manually annotated net-like structure of the microrelief on skin.

minimizes the energy of the form [89]:

$$\mathcal{M}(T, S \circ W) + \mathcal{R}(W) \tag{2.1}$$

where $\mathcal{M}$ quantifies the agreement between T and S, and $\mathcal{R}$ is the regularization term that favors user-defined specific properties in the transformation and helps to assuage difficulties arising from the ill-posedness of the registration problem. Transformation W is a mapping function of the domain Ω to itself; it maps locations in S to T. Specifically, for every point $\mathbf{x} \in \Omega$, W defines a displacement field $\mathbf{u}$ such that $W(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$.

There are three components in an image registration algorithm: deformation model, objective function and optimization method. [89] The deformation model defines the parameter space and inherently incorporates regularization so as to inform what transformations are acceptable and include prior knowledge, thus greatly helping to limit the solution. The optimization method navigates the parameter space of the often ill-posed problem with a non-convex objective function to find the local minimum. The ill-posed nature of the registration problem is due to the number of unknowns being greater than the number of constraints, as a displacement vector is estimated for the scalar information (i.e. image intensity) at every location.

The rest of this section discusses image correspondence from skin appearance and related work in skin registration and related fields.

2.3.1 Skin Appearance and Correspondence

Two general approaches exist for constructing the matching criteria $\mathcal{M}$ between target image T and source image S: (1) iconic method and (2) geometric method, using Sotiras' terminology [89]. Iconic methods evaluate image alignment based on intensity-based or information theoretic-based measures using correlation-like methods. According to [110], this approach has the advantage of better representing the dense correspondence between images accurately and does not rely on the presence of high-frequency details, thus it is widely used in medical imaging. On the other hand, the skin surface lacks high-contrast boundaries and has low-contrast glyphic patterns that appears to be repeated across the image. Thus, these attributes of the skin surface lead to a highly nonconvex optimization problem such that gradient-based optimization algorithms blindly converges to local minima far from the true solution in most cases.

The geometric method assesses image alignment by creating a matching criteria based on landmarks found in images. To find landmarks, first, a keypoint detector [70] is applied to image T and image S to obtain 2 sets of landmarks $\Psi = \{\psi_i \mid \psi_i \in \Omega, i = 1, \ldots, n\}$ and $\Lambda = \{\lambda_j \mid \lambda_j \in \Omega, j = 1, \ldots, m\}$ respectively. Unknown variables in the registration problem comprise correspondences between landmarks and parameters of the deformation model. This approach is advantageous as it is robust to various initial conditions and large deformations. After correspondences between landmarks have been found, fitting a deformation model to the landmarks is a relatively straightforward process. However, as only a sparse set of correspondences is used, the accuracy of interpolation or approximation of the dense deformation field decreases as the distance between landmarks increases. Therefore, accurate geometric method-based registration certainly requires identifying a uniform and dense distribution of correctly matched landmarks.

Finding correspondence between keypoints is a crucial step in the geometric method. In this thesis, we attempt to register a pair of images ($\Omega \in \mathbb{R}^2$ for 2D images) of the skin taken in same area from different viewpoints. To find the key-

point matches, the assumption we make is that same locations in both images appear to have similar intensities at the scale of skin features. We shall investigate this assumption in the following paragraphs.

The complex optical properties of skin layers and surface microgeometry of skin features give rise to the skin taking on different appearances when it is viewed and illuminated from different directions [18]. We can easily observe this phenomenon from specular reflections in regions of the skin surface when viewing it from an arbitrary viewpoint. In computer graphics, one widely used model for rendering skin is the Bidirectional Reflectance Distribution Function (BRDF) [75]. Given a 2D surface, the BRDF gives the relation of the reflected radiance dL_r at a point x in the direction $\boldsymbol{\theta}_r$ to the irradiance dE_i from a light source in direction $\boldsymbol{\theta}_i$:

$$BRDF(\boldsymbol{x}, \boldsymbol{\theta}_r, \boldsymbol{\theta}_i) = \frac{dL_r(\boldsymbol{x}, \boldsymbol{\theta}_r)}{dE_i(\boldsymbol{x}, \boldsymbol{\theta}_i)} \qquad (2.2)$$

A visual reference in Figure 2-6 shows that the BRDF assumes that incident light hits the same point as reflected light, despite light travelling through the epidermis and dermis layers of the skin. Despite the fact that skin is a volume scatterer [66], the model gives an approximately correct representation of surface reflection of the skin [35]. To our advantage, Marschner et al. [66] found that the BRDF of human skin with respect to reflected angle is close to constant for a light source near normal incidence, as seen in the plot in Figure 2-7. Hence, in this case, skin approximates a Lambertian surface: a surface in which the reflected radiance of every point on the surface is the same, regardless of the observer's angle of view. Moreover, the imaging device described in Chapter 2.1 is designed to achieve constant illumination of the skin in the ROI. Due to constant illumination and a Lambertian scene, our skin images fulfill the conditions [99] needed to confidently establish local correspondences between both viewpoints of the scene.
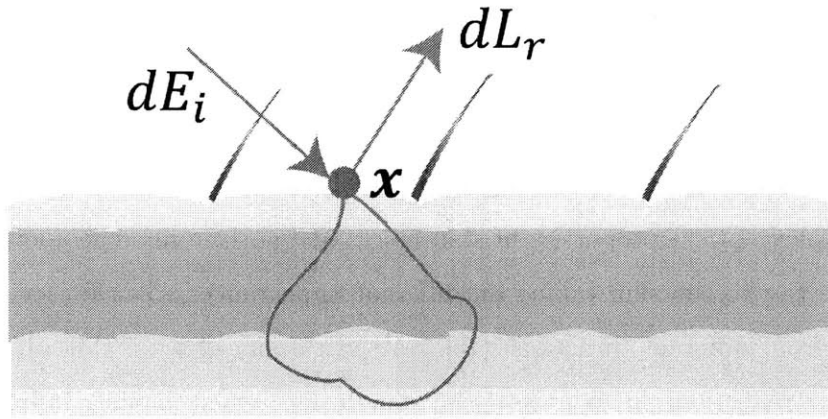
Figure 2-6: BRDF assumes that incident and reflected light enter and exit at the same point. When light enters the skin, some subsurface scattering occur. Image adapted from [35].
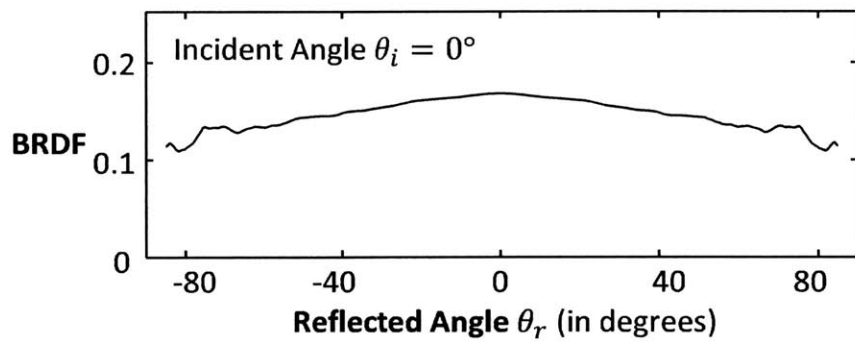
Figure 2-7: 1D BRDF at a single point of the human skin surface with respect to the angle of view when illumination normal to the surface is used. Image adapted from [66].

2.3.2 Related Work on Skin Registration

Change detection of individual PSLs can be used to diagnose diseases like melanoma [67]. Due to camera motion relative to the skin surface, accurate registration is desired for this application. However, in the commericial market, many computer-assisted diagnosis (CAD) systems for PSLs only offer the functionality of side-by-side or blink comparison of skin dermoscope images taken over periods of time [47]. In the literature, a few rigid registration methods have been developed specifically for change detection of PSLs. These methods assume that the correspondence between images is related by a global transformation. This assumption may hold for PSLs in dermoscope images because the skin surface is flattened into a plane by the dermoscope's glass plate or lens [27]. For instance, [86] registered two skin images with a PSL in the translation and rotation space using stochastic gradient descent to optimize mutual information. Also, [77] and [62] devised similar methods for estimating scaling and rotation parameters by a fast Fourier transform-based cross-correlation algorithm and translation parameters by hill-climbing optimization using non-parametric statistical similarity measures. These abovementioned iconic approach-based methods share the share similarity measures as some of the common techniques used in nonrigid registration of medical images [44]. But for skin surfaces without PSLs, the iconic approach may be unreliable, as explained earlier.

Another group of methods for rigid skin registration involves identifying a sparse set of landmarks in both images and aligning the images by a homography or affine transformation (i.e. the geometric method). These algorithms follow the well-known method proposed by Lowe [59]. The workflow for Lowe's method is as follows: First, keypoints are detected from each image. Such keypoints may be detected using for example, Harris corners [68], information theory [40] or Laplacian-of-Gaussian [59]. Second, a local descriptor around each keypoint is computed. For instance, some affine-invariant local descriptors [69] are based on image gradients [59], Haar wavelets [10], moment invariance [98] or shape contexts [6]. Third, correspondences between keypoints and the correct global transformation are found by employing different

31

variants of random sample consensus (RANSAC) [22]. These methods have been applied to constructing panoramas [9], multiview geometry matching [82] and fundamental matrix estimation [87]. One of the earliest use of the geometric method for registering skin images can be traced to Zambanini et al. [106], who used it to register skin images with cutaneous hemangioma, which are a benign red PSL on the skin developed in infancy due newly-formed blood vessels [19]. They employed Lowe's Scale Invariant Feature Transform (SIFT) [59] and RANSAC to calculate the optimal homography transformation. However, keypoints found were constrained within the region of the lesion after segmentation. Later, [4] used a similar method to register images of melanocytic nevus using keypoints found in the entire image domain. Next, Dong [54] demonstrated that keypoint matching via homography warping could be applied to matching skin images on faces by modifying SIFT to be more receptive to the shape of facial pores. This result was used to match human faces from multiple viewpoints using skin features. Lastly, Madan et al. [61, 60] showed that the same approach (using traditional SIFT) could be used to register multimodal images of facial skin by homography transformation. In this case, five imaging modalities of the skin were registered: optical images of the skin illuminated with unpolarized, parallel-polarized and cross-polarized light, and florescence images of the skin illuminated with ultraviolet and blue light.

Nonetheless, rigid registration does not allow us to accurately register the free skin surface. For example, homography transformation assumes that the scene is planar, hence the transformation does not model situations in which the skin is locally deformed or different viewpoints of a slightly curved surface like the forearm. Literature we have reviewed above shows that rigid registration is generally successful for a small, locally planar ROI or a fixed viewpoint shared between images. In contrast, for a freehand imaging system, a registration algorithm that can handle a larger viewpoint changes and a larger field of view is desirable. Thus, we may not assume a planar scene.

Nonrigid or deformable registration has been previously explored in skin-related literature to assess the mechanical properties of skin *in vivo*. Several studies have

proposed methods to find the full-field 2D displacement of the skin surface in response to stress. For example, in [64], an array of reflective markers was placed on the forearm and the displacement of the marker set was tracked with an optical imaging system, and the deformation field was interpolated. Using confocal laser scanning microscopy (CLSM) to reveal layers within the skin, [71] compared the results of a cubic B-spline free-form deformation (FFD) registration algorithm [80] with the mean squared error metric to that of the the Demon registration algorithm [94] for registering pairs stress-deformed and fixed 2D CLSM images. In comparison to existing works, our skin registration algorithm described in Chapter 3 is able to compute the deformation field of the skin from visible light images without physical markers. Furthermore, it can accurately register deformed skin images resulting from realistic viewpoint changes or small stretches in a 1 inch by 1 inch sized ROI.

2.4 Digital Hair Removal

In the literature, the primary use of DHR methods is to disocclude hair in dermoscopic images for automated analysis of PSLs. In these methods, the first step is to identify pixels in the image that contain hair and the second step is to replace these pixels with plausible skin colors based on neighboring skin colors and structure by image impainting. This implies that statistical estimates of missing data in hair-occluded regions are used instead of true, observed data. The major challenges lie in finding and replacing hairs that are thin, overlapping, or of similar contrast or color to underlying skin and microrelief structure. [46]

The key focus of our work in Chapter 5 is to address the first step of DHR: the image segmentation problem. Image segmentation is defined as the optimal partitioning of an image I into non-overlapping regions that are similar in characteristics such as texture or image intensity. Given image domain Ω, we wish to determine sets

$S_k \subseteq \Omega$ such that [78]:

$$\Omega = \bigcup_{k=1}^{K} S_k \qquad (2.3)$$

where $S_k \cap S_j = \emptyset$ and K is the number of regions. In DHR, we have $\Omega \in \mathbb{R}^2$ for pixels in 2D images, and $K = 2$ to represent the set of hair pixels and set of skin pixels. As convention, we assign binary labels of 0 to skin pixels and 1 to hair pixels in segmentation masks.

The rest of this section describes prior work on DHR in terms of hair segmentation and image inpainting methods.

2.4.1 Related Work on Digital Hair Removal

Hair Detection and Segmentation

In DHR, hair detection based on mathematical morphology [29] is widely used. Many of these methods are based on the closing-based top-hat transform. This morphological operation returns objects that are smaller than the structuring element and are darker than their surroundings. In 1997, Lee et al. [53] released DullRazor, the first and most well-known method for DHR, as a readily available software on Windows or Unix platforms [2]. The software uses generalized morphological closing operations with three structuring elements that model line orientations in horizontal, vertical and diagonal directions and then takes the maximum response from each pixel in the RGB image. Next, the difference between the intensity and response at each pixel is taken, and a predetermined threshold is applied to obtain a hair segmentation mask as the union of hair masks from each channel. Furthermore, Schmid-Saugeon et al. [83] used similar method to that of Lee et al. by applying generalized morphological closing operations with a disk structuring element in the luminance component of the CIELUV color space to obtain a hair mask. Moreover, Xie et al. [103] and Fiorese et al. [21] both used variants of the top-hat transform and hard thresholding to obtain a hair mask. The top-hat transform used in these methods is reportedly generally

34

effective at enhancing dark hairs with sufficient contrast on light skin, including both thin and thick hairs. In addition, to find light hairs on dark skin, the opening-based top-hat operator can be used.

Koehoorn et al. [46] presented a DHR approach that decompose the skin image into its luminance threshold set (256 binary image layers) and applies a multiscale morphological gap detection technique to find thin structures in each threshold layer. The layers are then merged to create a hair mask. The authors evaluated their algorithm to be effective at detecting many types of hair on different skin types. Due to the need to apply morphological operations on up to 512 binary threshold layers to detect dark and light hairs, the algorithm requires fast parallel processing techniques in order to process the skin image in a comparable time frame to other methods.

The next group of hair detection methods are edge detection-based. The authors of E-shaver [42] used the Radon transform to detect the predominant orientation of hairs and then used the Prewitt edge filter and thresholding to find light and/or thin hair edges in the predominant orientation. Tossi et al. [97] proposed a hair detection method using the adaptive canny edge detector. Lastly, Maglogiannis [63] compared edge detection results from combinations of Laplacian, Laplacian of Gaussian and Sobel methods. In these hair detection methods, the hair mask is obtained from the edge image by dilation with a fixed structuring element to fill holes in the interior regions of hair edges. However, we observed that this group of methods is prone to identifying non-curvilinear objects on the skin as hair because they assume that all high-contrast edges in skin images are belong to those of hair objects.

Matched filtering methods make up another group of methods used for hair detection. A 2D kernel, designed to model a the cross-sectional intensity profile of hair in the image at some unknown orientation and position, is convolved with the skin image. The resulting matched filter response indicates confidence in the prescence of a feature at each pixel. These methods make the assumptions [74] that the cross-section intensity profile of hair is approximated by a Gaussian (see Figure 2-8), and that hair has a small curvature and can be approximated by piecewise linear segments. In addition, the Gaussian profile may be inverted to extract dark or light

35

hairs. Abbas et al. [3] presented a matched filter with a 1D Gaussian profile and adaptive threshold based on the local mean of the 1D first derivative of Gaussian that helps to de-emphasize step edges that arise from a PSL in the image. Nguyen [74] used the normalized correlation of a 1D Gaussian profile, thresholding and skeletonization to find the centerlines of the hairs. Then, the hair mask is reconstructed based on the Gaussian cross-sectional profile of hairs. The normalized correlation (as opposed to convolution) helps to extract both light and dark hairs at the same time, but we found that the author introduced many parameters and steps in order distinguish both structures in a method that may not be robust enough to extract hairs from images with dense amounts of hair. Lastly, Huang et al. [33] found the maximum normalized correlation response using a multiscale 1D Gaussian profile and used hysteresis thresholding to obtain a hair mask. Thereafter, local linear discriminant analysis using the partial information of hair colors in the Lab color space and region growing recovered hair pixels at intersections. However, Koehoorn et al. [46] reported that this method is susceptible to producing dark halos where hairs were identified and removed. In general, matched filtering methods are able to produce smooth hair outlines and recover small gaps in the hair structure due to noise in hair pixels. But the skin image needs to be convolved with a large set of filters, which increases its computational cost. However, in practice, this process can be implemented in parallel. In our hair segmentation algorithm in Chapter 5, we introduce a multiscale second deriative of Gaussian matched filter and show that its response to hair objects is superior to many hair detection methods discussed so far.

Other methods to extract hair have also been used. Zhou et al. [109] used a curve fitting approach. They first enhanced dark hairs using generalized morphological closing operations with a disk structuring element and then used a differential geometric approach [90] to detect lines. The line segments are refined by Bézier curve fitting using RANSAC and a line intersection refinement procedure. Finally, the hair mask is created by dilating the skeleton according to the estimated width of each hair from the average normal intensity profile of the lines. Although the authors claim that the results are a clear improvement compared to the work of Schmid-Saugeon

36
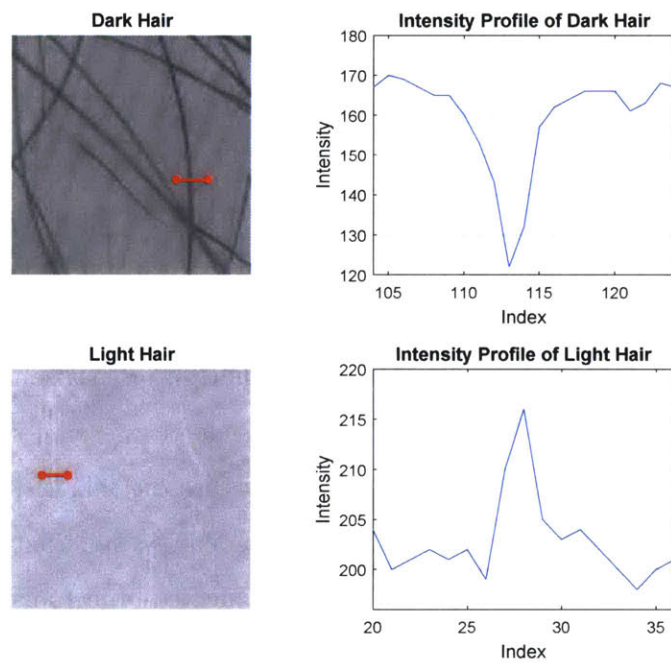
Figure 2-8: Intensity profile of hair showing the intensity profiles taken along the red lines in grayscale images of dark and light hair. In matched filtering, the intensity profile is assumed to be Gaussian.

et al. on 460 dermoscopy images, the results for this work are not well-reported and not reproducible due to lack of details in the algorithm description.

Xie et al. [104] proposed using an isotropic nonlinear filtering [57] to detect hairs as wide lines. They optimized their hair detection parameters to segment dense dark hairs on light skin that exhibit clumping in dermoscope images. To acquire the preliminary hair mask, an adaptive threshold based on linear regression of a linear combination of several global parameters is used. Then, pixels that are similar to the median of the skeleton image of the preliminary hair mask are marked as additional hair pixels. The drawback of this hair detector is that the isotropic nonlinear filter, however, also functions as a small object detector. Thus, the response appears noisy as it does not only detect curvilinear objects.

Typically, after obtaining a hair mask, many DHR methods use mathematical morphology to identify long and thin hair objects and remove non-hair objects. This step removes many false positives from the hair mask; the hair mask is usually peppered with short, disconnected, hair-like fragments of multiple orientations arising from skin features. Hence, removing these false positives help to preserve skin pigmentation, microrelief structure and texture. For example, Fiorese et al. [21] removes non-hair connected objects in the hair mask according to density, sphericity and convex hull sphericity metrics. Abbas et al. [3] filters out connected objects using morphological area opening and Haralick circularity [28] conditions. Xie et al. [104] uses circularity defined by the ratio between area and the minimum circumscribed circle radius of the connected regions to remove those regions. Lastly, Koehoorn et al. [46] discarded and refined objects from the hair skeleton mask: First, they discarded connected regions in the hair skeleton mask with small areas. Second, they pruned sours from the hair skeleton. Third, they removed objects according to junction-based metrics. Fourth, they constructed the hair mask by dilation using structuring elements of various radii.

Image Inpainting of Hair

Inpainting methods in existing DHR algorithms include simple bilinear interpolation [53, 74], replacement of pixels with pixels from the morphologically closed image

[83] and median filtering [33]. The above-mentioned methods do not attempt to preserve image structure and may lead to noticeable blurred appearance and color bleeding [109]. Other inpainting methods used include partial differential equation-based [21, 103], coherence transport [3], fast marching [45] and exemplar-based [109] methods. These methods reportedly model the texture of surrounding skin. However, we show in Chapter 6 that they do not preserve the true underlying skin texture. To overcome this shortcoming, our inpainting method records multiple skin images of the same area and combines information on underlying skin texture from the images in order to reconstruct a hair-free skin image with true skin texture.

2.5   Microrelief Junction Point Matching

Junction points of fingerprint patterns, or fingerprint minutiae, form the basis for many fingerprint matching methods [39, 37, 95, 108] due to the stability of these features over time. Similarly, we postulate that junction points of the microrelief structure of skin in general form a set of good candidate features that do not change over a period of a few years. In order to assess the stability of junction points, we need to be able to find matches between two sets of them. The problem of matching junction points is defined as finding the optimal matches between a reference point pattern set and target point pattern set in $\mathbb{R}^2$. To solve this problem, several challenges exist: First, different image acquisition parameters and the microrelief extraction method introduce many missing or spurious patterns to the point pattern sets. Second, the transformation is nonrigid due to the non-planar and stretchable nature of the skin surface. Third, there are as many as 7000 points in each set within a 1 inch by 1 inch patch of skin on the forearm. Thus, a well-known non-rigid PPM algorithm, thin-plate spline-robust point matching (TPS-RPM), with complexity $O(N^3)$ in the worst case, where N is the number of points in the set, would take a prohibitively long time to find matches. Hence, developing an efficient and robust non-rigid PPM algorithm to solve this problem is desirable. In Chapter 4, we present our nonrigid PPM algorithm and show that in preliminary experiments that it achieves low matching

error between points in pairs of skin image patches.

2.5.1 Related Work on Microrelief Extraction

Although it not the main focus of this thesis, microrelief extraction remains an integral part of MJPM. A microrelief extraction method takes a skin image as input and outputs the skeleton binary image of the microrelief structure, in a way similar to that seen in Figure 2-5. The junction points of the skeleton belong to the point pattern set.

In the literature, microrelief extraction is an intermediate step in algorithms used to quantify the degree of skin aging. Zou et al. [111], Kim et al. [43] and Choi et al. [15] employed the watershed transform [101] to segment the microrelief skeleton structure for further extraction of microrelief features such as mean cell area, line length or line width. In this thesis, we use a simple edge detection method based on local standard deviation to detect the microrelief structure.

2.5.2 Related Work on Non-Rigid Point Pattern Matching

Efficient PPM algorithms have been a topic of great interest in recent years. Recently, Du et al. [20] described a grid-based constellation PPM method that achieves fast, sub-pixel accurate and robust point matching for regularly-spaced points in a 2D grid. The authors noted that their method is faster and more robust to different local distortion scales and outlier points compared to established nonrigid PPM algorithms that utilize Gaussian mixture models (GMM) [73, 38] or thin-plate splines (TPS) [16]. The GMM method assigns a Gaussian component to each point in both reference and target point sets to construct two mixtures. Then, the GMMs are aligned by minimizing a statistical discrepancy measure between both GMMs. Du et al. commented that the GMM method is sensitive to a outliers and is easily trapped in local minima. The TPS method restricts the point sets to lie along curves, utilizing the TPS as the parameterization of the non-rigid spatial mapping. It also proposed soft assignment of correspondences. The final matches are determined through an

iterative expectation-minimization algorithm. Du et al. opined that the matching result was sensitive to parameter values. Additionally, they found that both methods were computationally intensive and slow. To dramatically improve speed, Du et al. used prior information such as the known correspondence at the bottom-left side of the grid, grid spacing and grid orientation. Correspondences are found by traversing the grid in a breadth-first search manner using the local radial nearest neighbor heuristic.

If a shape outline is drawn through all the grid points along the matching trajectory, and we can consider the work of Du et al. [20] as a simple shape or contour matching process using regularly-distributed landmarks. In Chapter 4, we extended their grid-based constellation PPM method to handle irregularly-spaced points of skin junction points by incorporating the shape context descriptor [5, 6], order-preserving assignment [84] and a novel concentric ring-based search space reduction.

Chapter 3

Skin Registration

Skin registration takes two images of skin patches and finds the optimal transformation between both images. In Chapter 2.3, we mentioned that skin registration techniques for visible light images have been mostly limited to rigid registration. Even if nonrigid registration is employed, physical markers placed on the skin surface were used. In this chapter, we present a procedure to accurately register deformed skin images resulting from a viewpoint change or stretching of an almost planar skin surface without the use of physical markers.

3.1 Algorithm Design

3.1.1 Overview

To map points $\mathbf{x} \in \mathbb{R}^2$ in source image S to points $\mathbf{x}' \in \mathbb{R}^2$ in target image T, we define a transformation $\mathbf{T} : \mathbf{x} \mapsto \mathbf{x}'$. First, to account for differing viewpoints in a scene and large-scale skin stretching, a global transformation $\mathbf{T}_{global}$ is applied. Next, to tackle deformation of the skin due to curvature of the surface and other nonlinearities, a local transformation $\mathbf{T}_{local}$ is applied. Hence, the combined transformation is:

$$\mathbf{x}' = \mathbf{T}(\mathbf{x}) = \mathbf{T}_{local}\big(\mathbf{T}_{global}(\mathbf{x})\big) \tag{3.1}$$

Figure 3-1 shows a flowchart of our skin registration procedure that takes a source
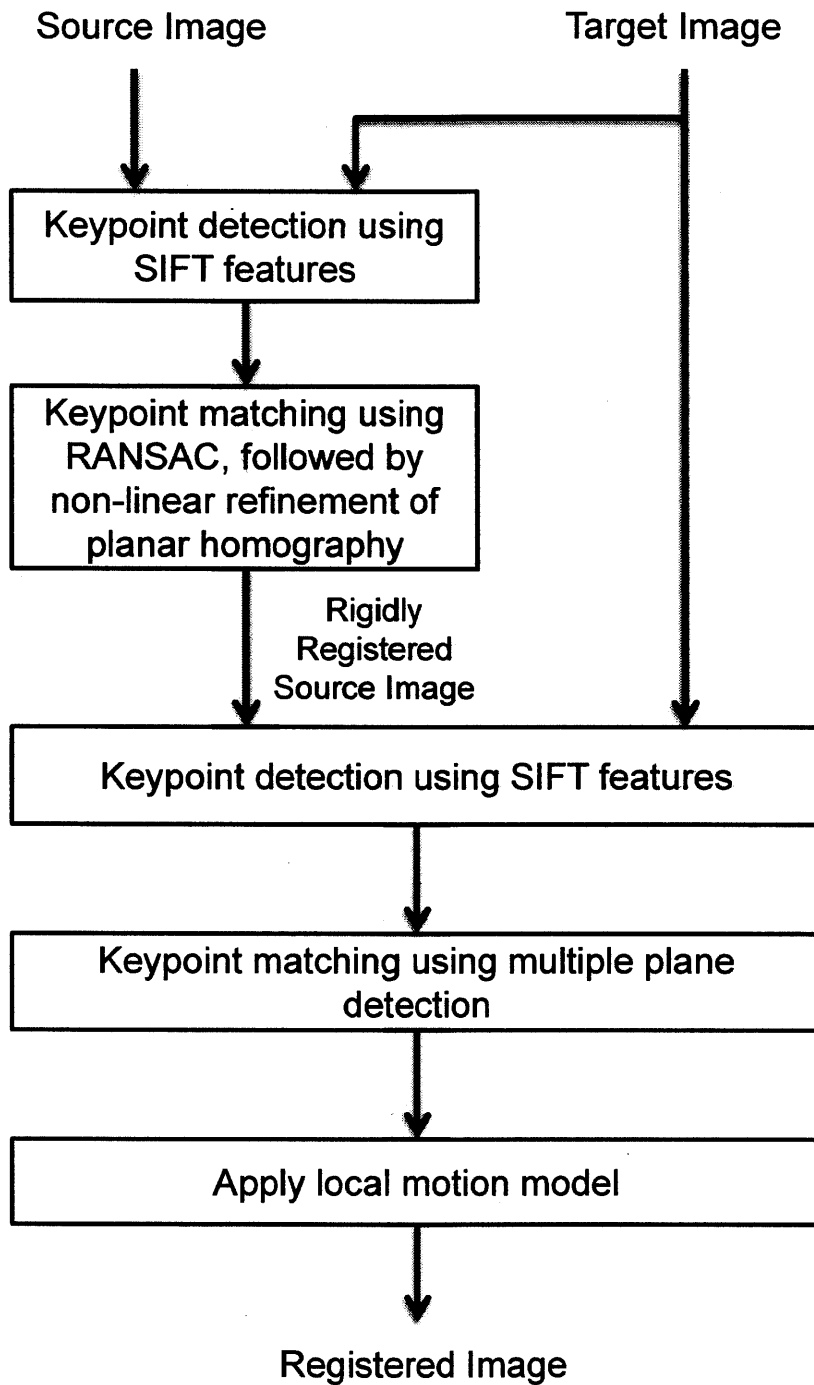
Figure 3-1: Algorithm flowchart for skin registration procedure.

image S and target image T as inputs and outputs the registered source image R. The first part of the procedure finds a global transformation by employing the commonly-used Lowe's method [59]: Scale Invariant Feature Transform (SIFT) keypoints are detected in S and T, then RANSAC estimates a homography transformation that warps S such that transformed source image S_H is "close" to target image T. In the second part, SIFT keypoints are detected in S_H. Then, the keypoints in S_H and T are matched using a multiple plane detection-based (MPD) method. In this work, we evaluate several of these MPD methods. Finally, a local motion model is applied to S_H to approximate or interpolate the unknown pixel locations in R based on the set of final matched keypoints K_f.

3.1.2 Keypoint Detection

Reliable landmarks or keypoints need to be identified in skin patches based on stable physical features of the skin. For this purpose, SIFT is used as the keypoint detector and feature descriptor, oweing to its success in detecting interest points on skin in various works [92, 60, 4]. A comprehensive description of SIFT is found in [59]. By observation, we note that SIFT identifies physical features mostly from the microrelief structure, but skin pigmentation and hair follicle openings are detected at the same time. In the following, we recommend changes in parameter values to enable SIFT to be more receptive to detecting skin features. SIFT detects the local extrema of multiscale difference-of-Gaussian (DOG) of a grayscale skin image. Because physical skin features are small, micro-level features relative to the image size, the DOG octaves are limited to the first 2 (i.e. original resolution and downsampling by 2). As skin features are of low contrast, the minimum contrast for the local extrema is set to 0.01 in order to accept any low contrast point. The threshold to eliminate edge responses remains at 10. In a typical skin image, thousands of keypoints are detected and they fully cover the entire image.

For a given location and scale, SIFT encodes the neighborhood of the corresponding keypoint using the histogram of gradient orientations at that scale. Rotation invariance is achieved by rotating the window according to the dominant orientation

of the histogram. The histogram is converted into the SIFT descriptor: a vector of 128 elements. In the first step, RGB source image S and target image T are converted into grayscale and the SIFT features are extracted to create two sets of keypoints.

3.1.3 Keypoint Matching and Global Registration

We define a 1 inch by 1 inch region of interest (ROI) at the center of the skin image. In our experiments, this ROI is approximately indicated by a red square inked onto the skin surface serving only as a visual reference for human observers. We assume that skin in the ROI is approximately on a flat 2D plane. So we model the global transformation between keypoints as a homography $T_{global} = \mathbf{H}$, which is a 3×3 invertible matrix that transforms between viewpoints in projective space:

$$s_H \hat{\mathbf{x}} = \mathbf{H} \mathbf{x} \tag{3.2}$$

where $\mathbf{x} = [x \ y \ 1]^T$ and $\hat{\mathbf{x}} = [\hat{x} \ \hat{y} \ 1]^T$ are homogeneous coordinates of a pair of corresponding keypoints in images S and T respectively, and s_H is a scale factor.

To estimate this transformation, first we match the sets of keypoints from S and T using approximate nearest neighbor search through a *k-d tree* [25] in Euclidean space on the 128-dimensional SIFT feature space. For efficiency, a bin size of 50 is used. The nearest neighbor ratio test (NNDR) given by Lowe [59] removes many false matches between nearest neighbors by taking the distances to the two nearest neighbors to a keypoint, denoted by d_1 and d_2 respectively, and accept matches such that:

$$\frac{d_1}{d_2} < T_{NNDR} \tag{3.3}$$

where we set threshold $T_{NNDR} = 0.7$. Thus, the set of initial matched keypoints is obtained. This is only an approximate matching of the sparse sets of keypoints. If the value of T_{NNDR} is decreased, the number of initial matched keypoints increases, but more processing time in subsequent stages of the algorithm is needed and more

false matches are present within the set of initial matched keypoints.

Next, an estimate of the homography is obtained by random sample consensus (RANSAC) [22]. During each RANSAC iteration, 4 random pairs of initial matched keypoints, termed a minimal sample set (MSS), are sampled by Gaussian sampling [41] as inputs for the normalized direct linear transformation (DLT) algorithm [31], which then outputs a homography hypothesis model. The MSS is termed as such because 4 pairs of keypoints, representing 8 degrees of freedom, is minimum number of correspondences required to derive a homography hypothesis model. As the model is proposed from the MSS, we can also call this model the minimal sample model. The set of paired keypoints that fit within a threshold is called the consensus set (CS) of the model. The threshold is known as the RANSAC inlier threshold ζ, which decides how closely points must fit the model. The error of a pair of matched keypoints with respect to a homography model $\tilde{\mathbf{H}}$ is given by $R(\tilde{\mathbf{H}}, \mathbf{x}, \hat{\mathbf{x}}) = \|\tilde{\mathbf{H}}\mathbf{x} - \hat{\mathbf{x}}\|_2$. Formally, the CS is defined as [24]:

$$\mathrm{CS}(\tilde{\mathbf{H}}, D, \zeta) = \{(\mathbf{x}, \hat{\mathbf{x}}) \in D \mid R(\tilde{\mathbf{H}}, \mathbf{x}, \hat{\mathbf{x}}) < \zeta\} \tag{3.4}$$

where D is the set of initial keypoint correspondences. This process is repeated for M trials or RANSAC iterations. Since we wish to find the homography hypothesis model that maximizes the cardinality of the CS, denoted by $\tilde{\mathbf{H}}_{best}$, the model with the largest number of inlier matched keypoints K_{best} is stored each round. We detail the RANSAC algorithm [112] in Figure 3-2.

After all RANSAC iterations are completed, the estimated homography is the homography found from the normalized DLT algorithm while considering all inlier matched keypoints in the largest CS. Thereafter, the estimated homography and its corresponding inlier matched keypoints are inputs for the nonlinear iterative refinement of the homography parameters using the Levenberg-Marquardt algorithm [72], in which the sum of the Sampson error [17] of the inlier keypoints is minimized. A detailed description of this well-known nonlinear iterative refinement procedure is found in [31].

```
Input  : D, M, ζ
Output: H̃_best, K_best
1  MaxCSCardinality = 0
   // RANSAC iterations
2  for j = 1 to M do
3  |   H̃ = DLT(SamplePoints(D))
4  |   CS = {(x, x̂) ∈ D | R(H̃, x, x̂) < ζ}
5  |   if |CS| > MaxCSCardinality then
6  |   |   H̃_best = H̃
7  |   |   K_best = CS
8  |   |   MaxCSCardinality = |CS|
9  |   end
10 end
```

Figure 3-2: RANSAC algorithm for keypoint matching.

Finally, using Equation 3.2, pixel locations in the source image S are warped to obtain pixel locations in globally registered image S_H. The RGB pixel intensities in S_H are calculated by standard bicubic interpolation.

Following global registration, distortions due to viewpoint changes or stretching of the skin surface are corrected to a large extent. Hence, true pixel locations in S and S_H are close approximations of each other. However, local levels of distortions are obvious in interpolated areas of S_H that are spatially far away from any matched keypoint from rigid registration.

In the following sections, we explain our rationale for the sampling technique used and describe the DLT algorithm in detail.

RANSAC and Uniform Sampling

RANSAC relies on obtaining an MSS, which is the set of the minimal number of points required to estimate a model. We denote this minimal number of points as b. To estimate a homography model (8 degrees of freedom), $b = 4$ pairs of keypoints are required. If we denote the total number of initial paired keypoints as N_k, then the total number of searchable models is $\binom{N_k}{b}$, which becomes large for large N_k. Say $N_k = 5000$, a typical number of initial paired keypoints for skin registration, then $\binom{5000}{4} \approx 2.6 \times 10^{13}$. Fortunately, a large number of minimal sample models correspond

to ground truth models, so we need only look at $M \ll \binom{L}{b}$ models. While there is no guarantee of success, we can set a sufficiently high M to find at least one ground truth model with high probability. Suppose that the probability of drawing an inlier for a particular model is p, then the probability of success P for obtaining a ground truth MSS by uniform, independent sampling is:

$$P \approx 1 - 1(1 - p^b)^M \tag{3.5}$$

which can be rearranged into:

$$M \approx \frac{\ln(1 - P)}{\ln(1 - p^b)} \tag{3.6}$$

Using Equation 3.6, we plot M against p for different P, which shows that given p is not too small, the number of MSS's required is tractable. For instance, given the empirically found $p = 0.4$ for skin images used by Sun [92], a simple calculation for $P = 0.99$ shows that the number of trial models M needed is only 178. On the other hand, suppose that $p < 0.2$ in noisy skin images, then M grows exponentially and computation time for RANSAC increases drastically.
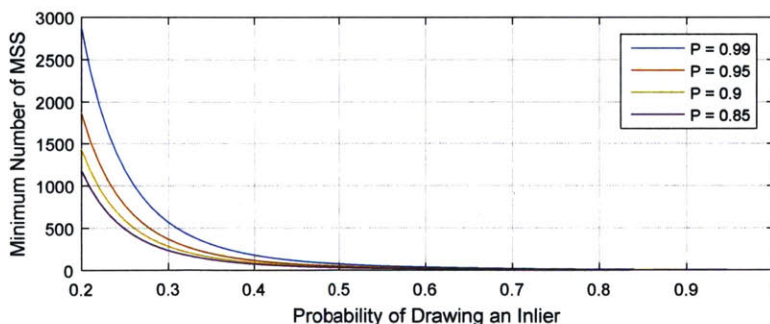
Figure 3-3: Plot showing the minimum required number of models M against the probability of drawing an inlier p using uniform sampling for various probabilities of success P.

Gaussian Sampling

To reduce the minimum required number of iterations to reliably detect a ground truth model, nonuniform sampling based on prior knowledge may be used. In a skin image, our general observation is that coplanar keypoints are more likely to be found in close proximity. Thus, there is a higher probability that an inlier keypoint is close to another inlier keypoint than one that is further away. We implemented a Gaussian sampling procedure described by Kanazawa et al. [41]. After an initial point $\mathbf{x}_i$ is drawn from a uniform distribution, a subsequent point selected $\mathbf{x}_j$ has the following probability of being drawn:

$$
P(\mathbf{x}_j | \mathbf{x}_i) = \begin{cases} \frac{1}{Z} \exp - \frac{\|\mathbf{x}_j - \mathbf{x}_i\|^2}{\sigma^2} & \text{if } \mathbf{x}_j \neq \mathbf{x}_i \\ 0 & \text{if } \mathbf{x}_j = \mathbf{x}_i \end{cases} \tag{3.7}
$$

where σ is the spatial scale of sampling and $Z = \sum_j P(\mathbf{x}_j | \mathbf{x}_i)$ is the normalization constant. σ is chosen by the user. If σ is too large, the sampling becomes uniform, while if σ is too low, only close neighbors of the initial point are sampled. We empirically found $\sigma = 2$ to strike a good balance. An analysis of Gaussian sampling on the minimum number of outlier-free MSS's required is provided by Toldo et al. in [96]. Their estimate is derived using additional parameters such as average inlier-inlier distance and inlier-outlier distance, which are difficult to ascertain for different data. Overall, their analysis shows support for using Gaussian sampling based on our assumptions.

Direct Linear Transformation Algorithm

The DLT algorithm [31] finds the linear least-square solution to $\mathbf{A}\mathbf{h} = \mathbf{0}$, where $\mathbf{H}$ in Equation 3.2 is rewritten as a vector $\mathbf{h} = [h_{11} \ h_{12} \ h_{13} \ h_{21} \ h_{22} \ h_{23} \ h_{31} \ h_{32} \ h_{33}]^T$ and $\mathbf{A}$

is a $2N \times 9$ matrix given by:

$$A = \begin{bmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1\hat{x}_1 & -y_1\hat{x}_1 & -\hat{x}_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1\hat{y}_1 & -y_1\hat{y}_1 & -\hat{y}_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N & y_N & 1 & 0 & 0 & 0 & -x_N\hat{x}_N & -y_N\hat{x}_N & -\hat{x}_N \\ 0 & 0 & 0 & x_N & y_N & 1 & -x_N\hat{y}_N & -y_N\hat{y}_N & -\hat{y}_N \end{bmatrix} \tag{3.8}$$

where x_i, y_i, $\hat{x}_i$ and $\hat{y}_i$ for $i = 1, \ldots, N$ denote coordinates of normalized paired keypoints and N is the number of keypoints considered. Normalization of coordinates, which improves the accuracy of the algorithm, is detailed in [32]. By setting $\|\mathbf{h}\| = 1$, the solution $\mathbf{h}$ is the unit eigenvector of the matrix $\mathbf{A}^T\mathbf{A}$.

3.1.4 Fast Initial Keypoint Matching

In the second part of skin registration, SIFT features are extracted from the transformed source image S_H and the set of initial matched keypoints from S_H are matched with the previous set obtained from source image S. Because the true pixel displacements between S_H and S are small, a grid-based method for fast initial keypoint matching is proposed. The grid-based method is inspired by the work in [51] used to find initial matches in high-resolution satellite image registration.

Here, we describe a fast grid-based initial keypoint matching method that limits the search space of the nearest neighbor search in NNDR while increasing the number of initial matches across the whole image pair domain. First, the spatial domains of S and S_H are divided into 2D tiles of size $T_{size} \times T_{size}$ (assuming square images and tiles). Corresponding tiles T_{ij}^S from S and $T_{ij}^{S_H}$ from S_H contain detected keypoints to be matched, where i and j are tile indices. To simplify notation, a tile is denoted by T_{ij}. We set $T_{size} = 100$ so as to typically obtain 100 keypoints in each tile for a typical skin image (see Figure 3-4). This provides a good compromise between poor match quality resulting from searching for matches among too few keypoints in a small tile and poor computational speed resulting from searching for matches among

too many keypoints in a large tile. For each i and j, the NNDR test (Equation 3.3) described in the first part of the skin registration procedure is applied to the sets of keypoints in the corresponding tiles. Note that the grid-based NNDR accepts more initial keypoint matches because the second nearest neighbor distance d_2 within a tile is greater or equal to d_2 found by the searching the entire image domain. This effectively increases the number of matched keypoints in the final keypoint matching result. Finally, a limit is set on the Euclidean distance between pairs of matched keypoints. If the distance between matched keypoints is more than 20, that pair of keypoints is removed from the list of initial matched keypoints. This threshold prevents unrealistically large local distortions due to outlier keypoints in the final registration result.
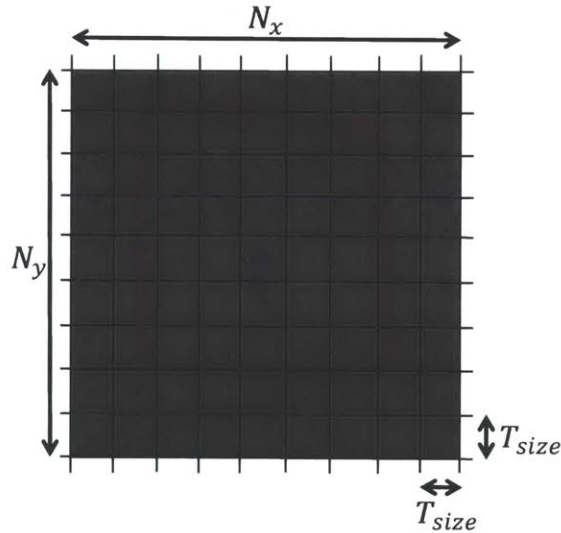
Figure 3-4: Tiles in skin image of size $N_x \times N_y$ having tiles of size $T_{size} \times T_{size}$.

3.1.5   Keypoint Matching by Multiple Plane Detection

The set of initial keypoint matches K_p (Figure 3-5a) contains many outliers because strict spatial relations are not enforced. Due to the close-to-planar skin surface scene, we propose a model in which the skin surface geometry is made up of several arbitrary overlapping 2D planes. If we find that a sufficiently high number of initial matched keypoints support a certain planar model, then there is a good chance that those

keypoints are matched accurately. The union of sets of matched keypoints from multiple planar models would provide a uniform and sufficently dense distribution of landmarks needed for the local motion model to approximate the dense deformation field accurately. This motivates discussion about how these planes can be reliably found. From the first part of our skin registration method, RANSAC identifies a single plane by homography, denoted by Π_1, and its set of corresponding keypoints K_1 in the skin patch scene (Figure 3-5b). Say that identified keypoints K_1 is removed from the set of initial matched keypoints K_p (i.e. $K_p - K_1$). Then the result, shown in Figure 3-5c, reveals that the density of correspondences in the area of plane Π_1 is significantly reduced, while other initial matched keypoints remain. It becomes obvious that it may be possible to apply a second round of RANSAC on the remaining keypoints to identify another plane Π_2 with its corresponding set of matched keypoints K_2. This simple process leads to the Sequential RANSAC algorithm — the repeated application of RANSAC to detect multiple models. The output of the process is the final set of matched keypoints, K_f.

<div align="center">(a)   (b)   (c)</div>
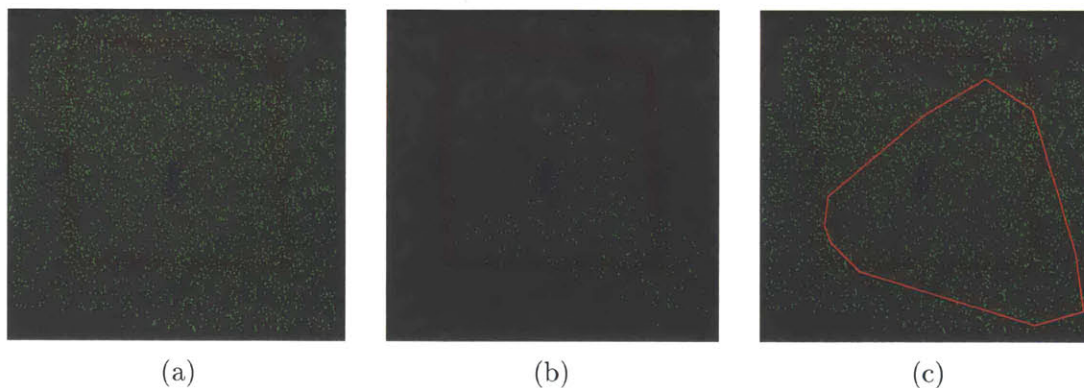
Figure 3-5: Process showing the set of initial matched keypoints remaining after removal of keypoints in the first plane Π_1 identified by RANSAC. Green points on the plot represent keypoints in the set of initial matched keypoints in the domain of target image T. (a) Set of initial matched keypoints K_p. (b) Set of matched keypoints in the first plane K_1. (c) Set of initial matched keypoints after removal of matched keypoints from the first plane $K_p - K_1$. The red outline shows an approximation of the enclosed region of Π_1 by taking the convex hull of keypoints in K_1. Density of keypoints in the enclosed region is shown to be reduced.

In the following sections, we describe several MPD methods, which are also termed

outlier-robust multiple model estimation methods as their application is not limited to finding planes in a scene but also to finding circles or lines in 2D space [24]. These methods fit multiple instances of a model to data corrupted by noise and outliers. Our usage of these methods is not conventional because our skin patch scene does not contain easily identifiable, distinct planes; in contrast, MPD methods are usually applied on scenes with buildings and walls [100, 41]. Nevertheless, we show in our evaluation that this approach helps find outlier-free correspondences between keypoints on a close-to-planar scene.

Sequential RANSAC

Sequential RANSAC has been used for plane detection [100, 41] between keypoints in pairs of images containing buildings or objects with large planar faces. It takes the set of initial matched keypoints K_p, number of randomly sampled minimal sample sets M_{SR}, the lower bound on the CS size N_{low} and RANSAC inlier threshold ζ as inputs, and outputs a set of sets of keypoints, denoted by $\mathcal{K} = \{K_1, K_2, \ldots\}$, each belonging to an identified plane in the scene (each set of keypoints is also termed a CS for a particular planar model). Matched keypoints belonging to planes are identified with RANSAC (Figure 3-2) sequentially and these matched keypoints are subsequently removed so that each set of keypoints is disjoint. The while- loop stops when the number of keypoints in a model is lower than N_{low}. Hence, the number of planar models found is dependent on the data and is not defined by the user. The algorithm in pseudocode is seen in Figure 3-6.

MultiRANSAC

In [113], Zuliani et al. noted that Sequential RANSAC has a drawback in that inaccurate model estimation in one round affects the estimates in later rounds. This is due to the greedy approach of maximizing the CS each round, which does not guarantee that ground truth models are found. And errors in model estimation is only made worse by the sequential approach. Therefore, Zuliani et al. introduced a parallel method of detecting models called MultiRANSAC [113]. MultiRANSAC

54

```
Input    : $K_p$, $M_{SR}$, $N_{low}$ , $\zeta$
Output: $\mathcal{K} = \{K_1, K_2, \ldots\}$
1  $\mathcal{K} = \emptyset$; $w = 1$;
2  while true do
3  |    $K_w = \text{RANSAC}(K_p, M_{SR}, \zeta)$
4  |    if $|K_w| \geq N_{low}$ then
5  |    |    $K_p = K_p - K_w$
6  |    |    $\mathcal{K} = \mathcal{K} \cup K_w$
7  |    else
8  |    |    break
9  |    end
10 end
```

Figure 3-6: Sequential RANSAC algorithm for keypoint matching.

looks for the best set of W models by updating every iteration the current set of W models with W proposed models, using a fusion procedure to merge both sets. The algorithm enforces all sets of keypoints after fusion to be disjoint so that W distinct models are identified. Thus, the number of ground truth models in the data must be known beforehand.

The pseudocode MultiRANSAC algorithm is described in Figures 3-7 and 3-8. Figure 3-7 describes the fusion procedure to combine two sets of sets of CS: the current set $\mathcal{K}_{current}$ and a proposed set $\mathcal{K}_{proposed}$, each with W sets of CS. First, the $\mathcal{K}_{current}$ and $\mathcal{K}_{proposed}$ are made into a combined set $\mathcal{K}_{combined}$ with $2W$ sets of CS. Then, $\mathcal{K}_{combined}$ is sorted according to decreasing order of cardinality of their CS. A merged set $\mathcal{K}_{next}$ with W sets of CS is constructed subject to each successive set of CS added to $\mathcal{K}_{next}$ being disjoint. Thus, the largest W sets of CS that are all disjoint with one another are retained in $\mathcal{K}_{next}$. Figure 3-8 describes the main body of MultiRANSAC. For each of the M_{MR} iterations, W disjoint MSS's are drawn from the set of initial matched keypoints K_p to produce the proposed set $\mathcal{K}_{proposed}$ of W CS. Then, UPDATECS merges $\mathcal{K}_{proposed}$ with the current best set $\mathcal{K}_{current}$.

There exists an automatic way to determine number of iterations M_{MR} in [113], based on the probability of success of identifying the ground truth models, which is derived from the cardinality of each CS in $\mathcal{K}_{current}$. However, in our evaluation, we simply choose this quantity to be user-defined because we are more concerned with

evaluating the accuracy of matched keypoints identified by this algorithm rather than minimizing time taken by the algorithm.

Input : $\mathcal{K}_{current}$, $\mathcal{K}_{proposed}$, W
Output: $\mathcal{K}_{next}$

1   $\mathcal{K}_{next} = \emptyset$; $\mathcal{K}_{combined} = \mathcal{K}_{current} \cup \mathcal{K}_{proposed}$
2   Sort $K_i \in \mathcal{K}_{combined}$ in order of decreasing cardinality
3   **for** $i = 1$ *to* $|\mathcal{K}_{combined}|$ **do**
       `// Stop when` W `models in` $\mathcal{K}_{next}$
4   **if** $|\mathcal{K}_{next}| == W$ **then**
5 | **break**
6   **end**
       `// Add` K_i `to` $\mathcal{K}_{next}$ `only if disjoint to all` K `in` $\mathcal{K}_{next}$
7   **if** $K_i \cap K == \emptyset$ *for all* $K \in \mathcal{K}_{next}$ **then**
8 | $\mathcal{K}_{next} = \mathcal{K}_{next} \cup K_i$
9   **end**
10 **end**
    `// If fail to find W models, return the best-so-far`
11 **if** $|\mathcal{K}_{next}| < W$ **then**
12 | $\mathcal{K}_{next} = \mathcal{K}_{current}$
13 **end**

Figure 3-7: UpdateCS routine in MultiRANSAC algorithm.

J-Linkage

Sequential RANSAC and MultiRANSAC both use CS to find models. A major shortcoming of using CS is that each data is assumed to belong to one model only because the disjointness constraint between models, but in actuality, the data can belong to two or more models. As a result, these algorithms may output incorrect models. A detailed discussion of this problem may be found in [24]. Thus, recent research in outlier-robust multiple model estimation have moved on to other criteria. Some of these methods include histogram of residuals [107], Mercer kernel fitting [14] and graph cuts [36] approaches. In this work, we describe and evaluate J-linkage by Toldo and Fusiello [96].

The criteria for J-linkage is the preference set (PS). Instead of creating sets containing data points that fit the models, sets containing models that fit the data points

Input : K_p, M_{MR}, W, ζ
Output: $\mathcal{K} = \{K_1, K_2, \ldots\}$

1 $\mathcal{K}_{current} = \emptyset$
2 **for** $j = 1$ *to* M_{MR} **do**
3     $K_{valid} = K_p$; $\mathcal{K}_{proposed} = \emptyset$
      // Propose W sets of keypoints/CS
4   **for** $w = 1$ *to* W **do**
5        $\tilde{\mathbf{H}}_{best} = \text{DLT}(\text{SamplePoints}(K_{valid}))$
6        $K_w = \{(\mathbf{x}, \hat{\mathbf{x}}) \in K_p \mid R(\tilde{\mathbf{H}}_{best}, \mathbf{x}, \hat{\mathbf{x}}) < \zeta\}$ (Equation 3.4)
7        $K_{valid} = K_{valid} - K_w$
8        $\mathcal{K}_{proposed} = \mathcal{K}_{proposed} \cup K_w$
9   **end**
10   **if** $\mathcal{K}_{current} \neq \emptyset$ **then**
11       $\mathcal{K}_{current} = \text{UpdateCS}(\mathcal{K}_{current}, \mathcal{K}_{proposed}, W)$
12   **else**
13       $\mathcal{K}_{current} = \mathcal{K}_{proposed}$
14   **end**
15 **end**
16 $\mathcal{K} = \mathcal{K}_{current}$

Figure 3-8: MultiRANSAC algorithm for keypoint matching.

are created. In our case, for each pair of initial matched keypoints $(\mathbf{x}, \hat{\mathbf{x}}) \in K_p$, we have:

$$\text{PS}(\mathbf{x}, \hat{\mathbf{x}}, \tilde{\mathcal{H}}, \zeta) = \{\tilde{\mathbf{H}} \in \tilde{\mathcal{H}} \mid R(\tilde{\mathbf{H}}, \mathbf{x}, \hat{\mathbf{x}}) < \zeta\} \tag{3.9}$$

where $\tilde{\mathcal{H}}$ is the set of minimum sample models. The PS of a set of matched keypoints is further defined the intersection of their individual PS's:

$$\text{PS}(X, \hat{X}, \tilde{\mathcal{H}}, \zeta) = \bigcap_{(\mathbf{x}, \hat{\mathbf{x}}) \in (X, \hat{X})} \text{PS}(\mathbf{x}, \hat{\mathbf{x}}, \tilde{\mathcal{H}}, \zeta) \tag{3.10}$$

where $(X, \hat{X})$ is the set of matched keypoints. That is, the homography model hypotheses that the combined set of matched keypoints fit.

Thus, J-linkage represents each data point with the characteristic function of the set of models preferred by that point. Toldo and Fusiello describe taking M_J number of MSS's to obtain M_J minimum sample models $\tilde{\mathcal{H}} = \{\tilde{\mathbf{H}}_1, \ldots, \tilde{\mathbf{H}}_{M_J}\}$, and each point

takes on a value of 0 or 1 depending on model membership in the M_J-dimensional *conceptual space* $\{0, 1\}^{M_J}$. In this space, points are represented as M_J-dimensional vectors and clusters are formed for points sharing similar models. To search for these clusters, J-linkage uses an agglomerative clustering strategy based on a dissimilarity measure between PS's in pairs of clusters. The measure used is the Jaccard distance:

$$D_J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|} \tag{3.11}$$

where A and B are sets and $D_J \in [0, 1]$. A Jaccard distance of 0 indicates that the elements in both sets are identical ($A = B$) while a Jaccard distance of 1 indicates that the elements in both sets are disjoint ($A \cap B = \emptyset$).

Agglomerative clustering works as follows: Initially, M_J data points are assigned to their individual singleton clusters $1, \ldots, c_{M_J}$. The current set of clusters is denoted as C. Next, every iteration, the two clusters with the smallest Jaccard distance between their corresponding PS's are merged. In other words, clusters sharing similar models are merged first. When the Jaccard distance in all pairs of clusters equals 1, then the process stops as all pairs of clusters are have disjoint PS's. That is to say that each cluster contains at least one model that fits all of its points. Clusters with large memberships correspond to ground truth models while clusters with small memberships are outlier models. Clusters with small memberships are removed by a user-defined threshold η. A pseudocode description of J-linkage is shown in Figure 3-9.

3.1.6 Local Motion Model

A local motion model takes the final set of corresponding keypoints K_f as input landmarks so as to warp the rigidly registered source image S_H. The model interpolates or approximates the dense pixel locations in the rest of the image domain and provides regularization for local transformations on image S_H. After warping and bicubic interpolation of RGB values from S_H, the final registered image R is obtained.

Input : K_p, M_J, η, ζ

Output: $\mathcal{K} = \{K_1, K_2, \ldots\}$

```
// Assign each pair of matched keypoints to a cluster
```

1   $C = \{\{k\} \mid k \in K_p\}$

```
// Get M_J minimum sample models
```

2   $\tilde{\mathcal{H}} = \emptyset$

3   **for** $j = 1$ *to* M_J **do**

4      $\tilde{\mathbf{H}}_\mathbf{j} = \mathrm{DLT}(\textsc{SamplePoints}(K_p))$

5      $\tilde{\mathcal{H}} = \tilde{\mathcal{H}} \cup \tilde{\mathbf{H}}_\mathbf{j}$

6   **end**

7   **while** *true* **do**

```
        // Pick a pair of clusters with the smallest Jaccard distance
```

8      $\{A, B\} = \underset{\{A,B\}\in Pairs(Clusters)}{\arg\min} \; D_J(PS(A, \tilde{\mathcal{H}}, \zeta), PS(B, \tilde{\mathcal{H}}, \zeta))$

9      $d_{min} = \underset{\{A,B\}\in Pairs(Clusters)}{\min} D_J(PS(A, \tilde{\mathcal{H}}, \zeta), PS(B, \tilde{\mathcal{H}}, \zeta))$

10   **if** $d_{min} == 1$ **then**

11 | **break**

12   **end**

```
        // Replace pair of clusters with their union
```

13     $C = (C - \{A, B\}) \cup \{A \cup B\}$

14   **end**

15   $\mathcal{K} = \{K \in C \mid |K| \geq \eta\}$

Figure 3-9: J-linkage algorithm for keypoint matching. For convenience, we define arbitrary letters to represent matched keypoints $(\mathbf{x}, \hat{\mathbf{x}})$ in any set of matched keypoints. Also, note that A and B are sets of keypoints.

B-Spline

Among free-form deformation (FFD) models, B-spline models are some of the most commonly used in medical image registration [80, 49]. On the image domain $\Omega = \{\mathbf{x} = (x,y)|0 \leq x < N_x, 0 \leq y < N_y\}$ with image size $N_x \times N_y$, a rectangular grid $\Phi = K_x \times K_y$ of control points $\phi_{i,j}$ with uniform spacing δ is superimposed. The image is deformed by displacing $\phi_{i,j}$. The dense transformation $\mathbf{T}_{local}$ is written as the summation of tensor products of univariate cubic B-splines:

$$\mathbf{T}_{local}(\mathbf{x}) = \sum_{l=0}^{3} \sum_{m=0}^{3} B_l(\mu_x) B_m(\mu_y) \phi_{i+l,j+m} \tag{3.12}$$

where $i = \lfloor x/N_x \rfloor - 1$, $j = \lfloor y/N_y \rfloor - 1$, $\mu_x = x/N_x - \lfloor x/N_x \rfloor$ and $\mu_y = y/N_y - \lfloor y/N_y \rfloor$. B_l represents the l-th basis function of the B-spline, given by:

$$B_0(\mu) = (1 - \mu)^3/6 \tag{3.13}$$

$$B_1(\mu) = (3\mu^3 - 6\mu^2 + 4)/6 \tag{3.14}$$

$$B_2(\mu) = (-3\mu^3 + 3\mu^2 + 3\mu + 1)/6 \tag{3.15}$$

$$B_3(\mu) = \mu^3/6 \tag{3.16}$$

Despite the drawback that the transformation model does not guarantee that the topology is preserved, it provides smooth C^2 continuous transformations (i.e. first and second derivatives are continuous) and only requires few parameters to describe local deformations: a control point $\phi_{i,j}$ with limited support is displaced to describe transformations in its local neighborhood. Hence, the degree of locality of a deformation is dependent on the resolution of the control point grid. A coarse grid models large, global-level deformations, while a fine grid models smaller, local deformations. Hence, we use the multilevel B-spline approximation (MBA) approach by Lee et al. [52] in order to take advantage of differing grid resolutions in modeling different kinds of deformations. In this approach, a hierarchy of control point grids $\Phi^1, \ldots, \Phi^L$ at increasing resolutions is defined, such that spacing δ is halved at each increasing level

up to finest resolution at level L. For each level l, a local transformation $\mathbf{T}^l_{local}$ is computed using the B-spline approximation method detailed in [52] that reduces the residual between paired keypoints. The total effect is that the sum of each $\mathbf{T}^l_{local}$ defines the final transformation $\mathbf{T}_{local}$:

$$\mathbf{T}_{local}(\mathbf{x}) = \sum_{l=1}^{L} \mathbf{T}^l_{local}(\mathbf{x}) \tag{3.17}$$

3.2 Experimental Results

3.2.1 System and Parameters

The algorithms were implemented in MATLAB Release 2015b on a system with an Intel Xeon E3-1225 v3 3.2 GHz processor. In following sections, we evaluated and compared the results of the three MPD methods with both real and simulated data. Real data consists of three pairs of images recorded from the ventral forearm of a male, 23-year-old subject of South Asian descent. They are RGB images with resolution 1540×1540. The dataset comprises pairs of images of the skin with differing viewpoints, with and without an applied stretch of the skin and recorded approximately 5 months apart. The images were used for visual evaluation of the construction process of planar models by the multiple-plan methods and registration results. The simulated dataset consists of images created by distorting a single image from the real dataset. Simulated images were used to evaluate the MPD methods for accuracy, robustness to in-plane rotations, nonlinear distortions and image noise, and computation time with respect to number of initial matched keypoints.

We set the following default parameters for the three methods evaluated, including the same parameters for rigid registration:

- Rigid Registration: $M = 1200$, $\zeta = 1$

- Sequential RANSAC: $M_{SR} = 100$, $N_{low} = 20$, $\zeta = 1$

- MultiRANSAC: $M_{MR} = 100$, $W = 12$, $\zeta = 1$

61

- J-linkage: $M_J = 1200$, $\zeta = 1$

In addition, we limited the number of models W found by Sequential RANSAC to 12 by replacing the while-loop in the Sequential RANSAC with an appropriate for-loop (see Figure 3-6). Thus, with these parameters, each algorithm drew a maximum of 1200 MSS's from the set of initial matched keypoints. We found that this total number of MSS's is sufficient for high probability of successful registration with the MPD methods while taking below a minute to run for most cases. However, at times, on skin image pairs with relatively few initial matched keypoints, initial rigid registration was run a few times in order to obtain a successful registration. But this has no consequence on evaluation of MPD methods because they were evaluated on the resulting rigidly registered image.

For J-linkage, setting the cluster membership threshold η for different image pairs is tricky. Having few initial matched keypoints usually results in the need to set η low but having many initial matched keypoints require a high η to reduce outliers. We explain our choice of η for each image pair.

3.2.2 Real Images with Differing Viewpoints

Rigid Skin Registration

Figure 3-10a is the target image T and Figure 3-10b is the source image S. The images are taken within seconds of each other, with a change of camera viewpoint, as seen from their difference image (Figure 3-10c), showing an approximate translation in the y-direction. For any pair of grayscale images A and B, the difference image D for all pixels (x, y) is given by:

$$D(x,y) = |A(x,y) - B(x,y)| \qquad (3.18)$$

The difference image allows us discern the alignment between images where the images have very similar intensity values. Higher intensities in certain regions correlate with misalignment. Although illumination differences between images also causes intensity

differences, we assume the error is negligible for this image pair. The 1 inch by 1 inch red square outline and markings inked on the skin surface serve as visual aids to indicate the ROI for human observers. There are a few dark, long hairs that occlude the skin surface. After global registration, image S is transformed into globally registered image S_H (Figure 3-10e). 176 matched keypoints were identified during homography-based registration, as shown in Figure 3-10d, superimposed on image T. Image S_H has fairly accurate alignment (Figure 3-10e), especially in the local neighboring regions around the matched keypoints. Outside these regions, there is observable misalignment, as seen from higher intensities found in the difference image 3-10f. Therefore, rigid registration only manages to register images within a local planar region approximately demarcated by the matched keypoints. As evaluation of rigid registration is not the key focus of this chapter, we direct readers to Chapter 4.2.2 for a detailed evaluation of rigid skin registration.

Evaluation of Grid-Based Initial Keypoint Matching

Our grid-based NNDR initial keypoint matching method is evaluated based on image pair S_H and T and the results are compared with the NNDR test (Equation 3.3). We draw the initial matched keypoints for both methods over image T in Figure 3-11. It can be seen that the density of keypoints is higher for grid-based NNDR. As shown in Table 3.1, the number of initial matched keypoints is notably higher. Because it found a higher number of inlier matched keypoints, the grid-based NNDR method could potentially increase the number of final matched keypoints, and thus, potentially increase the registration accuracy of the local motion model. Furthermore, our grid-based NNDR method is 66 times faster than traditional NNDR. However, some of the time savings may be eroded if MPD keypoint matching needs to search through more initial matched keypoints.

Failure of Free-Form Deformation Methods

In our evaluation for accuracy, nonrigid registration methods would be required to accurately register the entire ROI approximately enclosed by the 1 inch by 1 inch
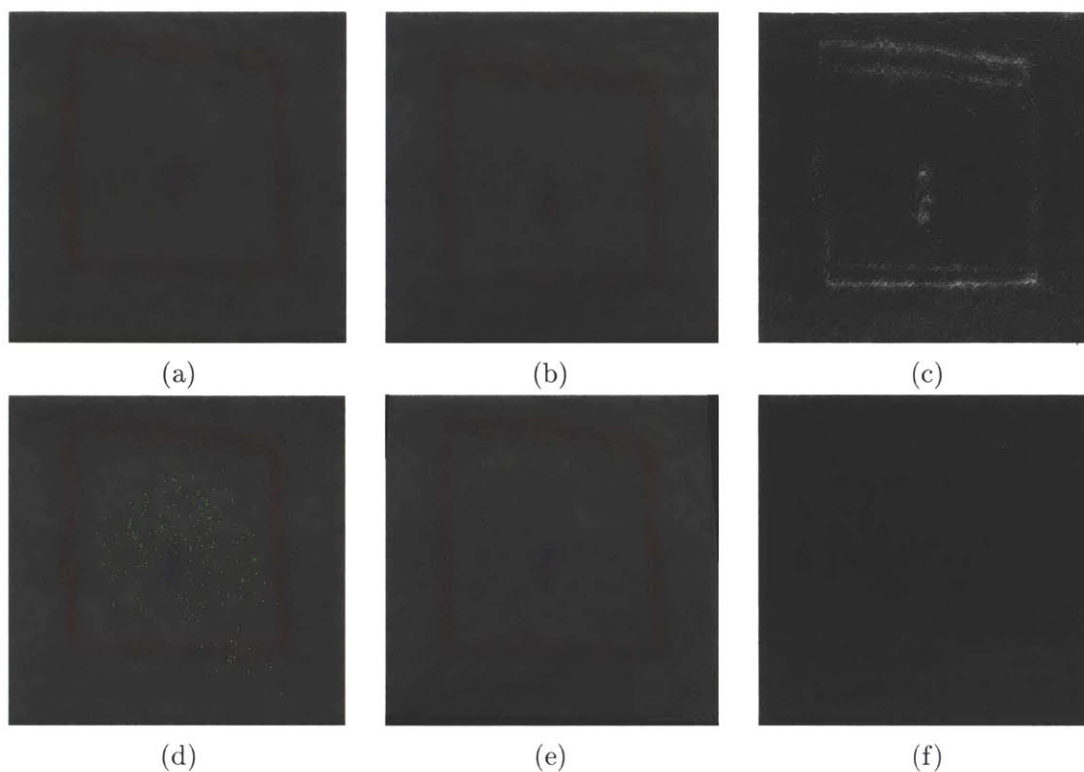
Figure 3-10: Images for registration and global registration results. (a) Target image T. (b) Source image S. (c) Difference image of T and S. (d) Corresponding keypoints (green markers) from global registration superimposed on T. (e) Globally registered image S_H. (f) Difference image of T and S_H.
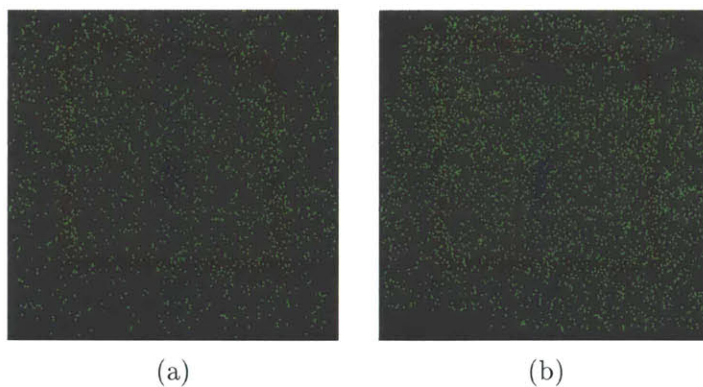
Figure 3-11: Comparison of initial keypoint matching results. (a) NNDR. (b) Grid-based NNDR.

Table 3.1: Comparison of number of initial correspondences and runtime for initial keypoint matching.

Method	Number of Initial Correspondences	Runtime (s)
NNDR	2087	10.59
Grid-Based NNDR	3552	0.16

red square outline. For comparison of our MPD methods with existing algorithms, we applied two well-known FFD methods on image pair S_H and T: a cubic B-spline FFD registration algorithm [80] with the mean squared error metric and Demon registration algorithm [94]. The registration results are displayed in Figure 3-12. The registered images R were highly erroneous; the texture of skin was not preserved and the visual markings on the skin did not resemble the shape seen in target image T. This demonstrates that registration based on the iconic method that evaluates image alignment based on measures on the entire image domain is ineffective for registering general skin images. These images lack large-scale structures like moles and possess the repetitive structure of microrelief, which makes these registration methods fail.

(a)   (b)
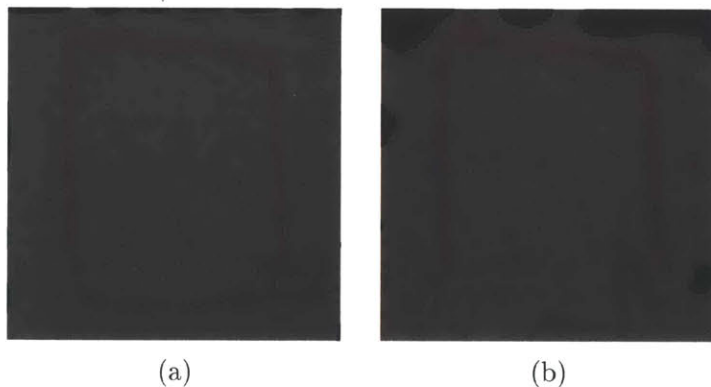
Figure 3-12: Failure of iconic deformable registration methods for registering two skin images. (a) Registered image for cubic B-spline FFD registration algorithm. (b) Registered image for Demon registration algorithm.

3.2.3 Comparing Multiple Plane Detection Methods

We now visually evaluate nonrigid registration using Sequential RANSAC, Multi-RANSAC and J-linkage methods on image pair S_H and T. First, matched keypoints

from the first 12 planar models identified from each method were plotted so that the model construction process could be evaluated. Second, registered images obtained from each method were compared with one another.

In Figure 3-15, matched keypoints from the $W = 12$ models for Sequential RANSAC are shown superimposed on image T. Each model contained at least 80 matched keypoints. Thus, in the figure, we observed that keypoints in each plane gave rise to locally dense clusters that approximately demarcated the regions of each plane in image T. Thus, the keypoints shown in each model were observed to be spatially clustered. Strong evidence of spatial clustering of keypoints in planes using Sequential RANSAC lends support for using the Gaussian sampling procedure for sampling initial matched keypoints; given a keypoint, there is a high conditional probability of finding keypoints in the neighboring regions. This procedure enables reduction of the number of MSS's required to guess the ground truth models.

In Figure 3-16, matched keypoints from all $W = 12$ models for MultiRANSAC are displayed superimposed on T. The number of matched keypoints dropped off significantly after the first few models. The first 3 planes clearly exhibited dense clusters of keypoints with more than 100 keypoints with clear planar regions approximated. After 4-th plane, there were fewer than 100 points in subsequent planes and planar regions became ambiguous; it reached 22 points at the 10-th plane and only 5 points at the 12-th plane. Due to the low number of points per plane in the higher-numbered planes, the matched keypoints in these planes may be in fact outliers. In a skin scene where planes formed are ambiguous, having to set the number of models present in a scene is disadvantageous as some of these models may not be ground truth models due to low number of keypoints identified in those models. To improve the result of MultiRANSAC, it could be helpful to remove keypoints in these planes from the set of final matched keypoints K_f.

J-linkage clusters matched keypoints that are part of similar planar models. The number of points in each cluster found from keypoint matching is shown in 3-13. Due to ambiguity of planes in the skin scene, it is not surprising that that many plausible clusters were formed. Clusters with small memberships are considered outliers and

we chose the threshold on the minimum number of points in a cluster $\eta = 30$ to remove these small clusters, thus leaving 33 clusters or planar models in the final set. These planes have noticeably lower number of points as compared to those of Sequential RANSAC, although the total number of planes found is greater. We plot the first 12 models in Figure 3-17 and observed that the set of matched keypoints only formed ambiguous planar regions due to the sparsity of keypoints, unlike the case for Sequential RANSAC. In several planes, there were even a number of isolated keypoints, but it not clear whether these keypoints were outliers.
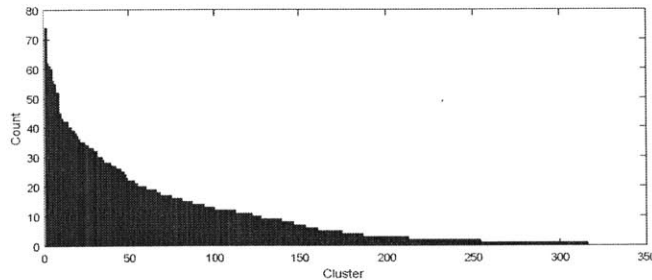
Figure 3-13: Number of keypoints in each cluster found by J-linkage.

Therefore, Sequential RANSAC displayed a clear advantage in the number of points found in each of its 12 planes as compared that of MultiRANSAC and J-linkage. This is desirable as a high number of keypoints in an identified model greatly increases the probability that the model is a ground truth model, and thus, increases the probability that all matched keypoints in the model are inliers. In general, repeated runs of MPD methods show that many plausible overlapping planar models exist due to the nondeterministic nature of random sampling. Random sampling gave rise to different sets of matched keypoints and planes each time an algorithm is run. Hence, the distribution of keypoints and number of keypoints in a plane were not the same each time. Thus, using MPD methods, it is difficult to obtain definitive results on the presence of outlier keypoints in a each plane. A simple visual test for detecting outliers in the set of final matched keypoints K_f is to look for unnatural local distortions in registered image R. These local distortions come about because the local motion model assumes that all final matched keypoints are inliers.

After taking the union of matched keypoints in all 12 planes for Sequential RANSAC,

the final set of matched keypoints K_f is obtained (Figure 3-14a). The 1761 matched keypoints in K_f formed a dense, evenly distributed cluster of landmarks covering the entire ROI. The local motion model used K_f to infer the nonrigid transformation; the registered image R is shown in 3-14b. R showed no signs of unnatural local distortions and appeared realistic. The difference image of R and T (Figure 3-14c) revealed that there was accurate image alignment in the entire ROI. The small differences in image intensities were mostly due to hair movement, different appearances of the microrelief structure at different viewpoints and out-of-focus regions at the extremities of the images. Therefore, Sequential RANSAC performed excellently in aligning images with different viewpoints.

By taking the union of matched keypoints in all 12 planes for MultiRANSAC, the final set of matched keypoints K_f is obtained (Figure 3-14d). Compared to that of Sequential RANSAC, there was a 46% decrease in the number of matched keypoints in the set of final keypoints K_f (936 points). The density of points was lower and the number of points was noticeably sparse at the top-right corner of the ROI, where there were more occluding hairs. The large reduction in number of final keypoints was due to the disjointedness constraint of matched keypoints in each planes and parallel procedure of MultiRANSAC. Hence, many inlier matched keypoints were removed from consideration. Although no unnatural distortions were seen in the registered image, The difference image of R and T (Figure 3-14f) showed that while there was accurate alignment in the middle of the ROI, the top-right corner of the ROI is slightly misaligned.

In Figure 3-14g, matched keypoints from all 33 models for J-linkage are shown superimposed on T. Like Sequential RANSAC, J-linkage revealed a dense, evenly distributed cluster of landmarks covering the entire ROI. There were 1381 keypoints in K_f for J-linkage. The registered image R is seen in 3-14h and no unnatural distortions were observed in the image. The difference image of R and T (Figure 3-14i) revealed that J-linkage managed to mostly accurately align the source image to the target image.

In this image pair with differing viewpoints, all three MPD methods performed

reasonably well for registration. No unnatural distortions in the registered image for the three methods, thus indicating that obvious outliers were not present. Sequential RANSAC performed the best as it identified the most final matched keypoints, while MultiRANSAC performed the worst because there are some degree of noticeable misalignment of between target and registered images.

3.2.4 Real Images of Stretched Skin

Our application for nonrigid skin registration is deducing the deformation field due to stress based on skin features alone. The deformation field can then be used to register the source image. We now demonstrate how well different MPD methods work on an image pair of skin with and without an artificially applied stretch. In Figure 3-18a, the target image T is shown. A stretch was applied upwards of T, causing the skin to deform, as shown in the source image S in Figure 3-18b. We superimposed both images on each other in Figure 3-18c. On superimposed images, the green image is the target image, while the red image is either the source or registered image. Next, rigid registration was applied. Figure 3-18d shows that 16 matched keypoints were found during rigid registration and the transformed image S_H is shown in Figure 3-18e. Figure 3-18f shows that most of deformation from stretch was already corrected by rigid registration, but a slight misalignment in many regions occurred across the ROI. During nonrigid registration, more keypoints were identified (Figures 3-18g, j and m). The registration result is shown in Figures 3-18h, k and n. The superimposed images in Figures 3-18i, l and o shows that all 3 MPD methods resulted in improved alignment in the top half of the image as compared to the rigid registration result. This is observed from the increased sharpness of the illustration. However, the bottom half of the image was misaligned. Due to stretching, the appearance of the microrelief changed at the scale of SIFT features used. In this region, the microrelief in image S appeared smoothed-out compared to that of image T, hence keypoints were not detected. Nonetheless, at the top region, the microrelief appeared compressed, but SIFT keypoint detection was able to detect the same keypoints. Therefore, a key limitation in using our skin registration technique to obtain the deformation field of
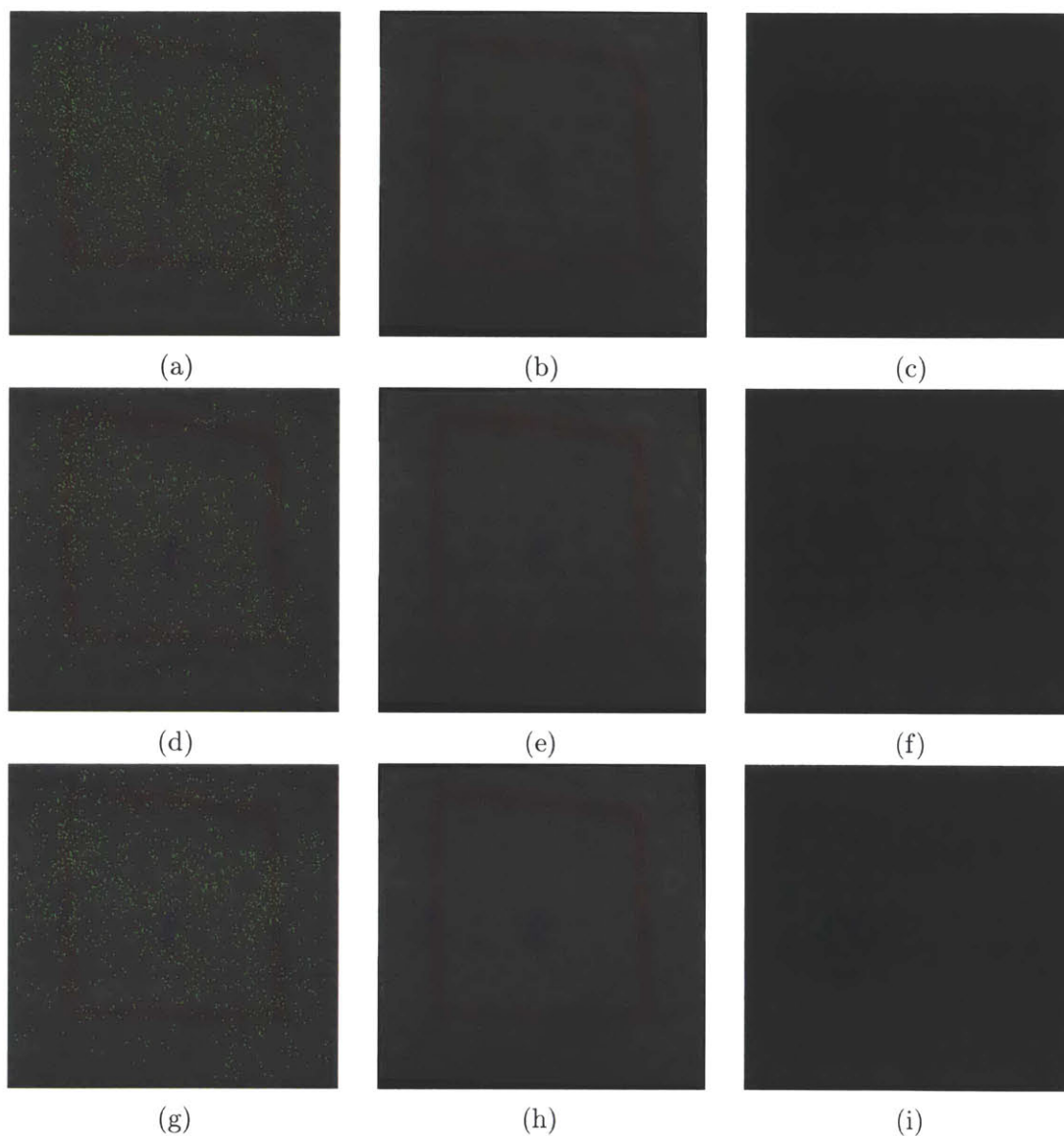
Figure 3-14: Registration result for differing viewpoints. (a-c) Sequential RANSAC: (a) Final set of 1761 matched keypoints K_f superimposed on T. (b) Registered image R. (c) Difference image of T and R. (d-f) MultiRANSAC: (d) Final set of 936 matched keypoints K_f superimposed on T. (e) Registered image R. (f) Difference image of T and R. (g-i) J-linkage: (g) Final set of 1381 matched keypoints K_f superimposed on T. (h) Registered image R. (i) Difference image of T and R.
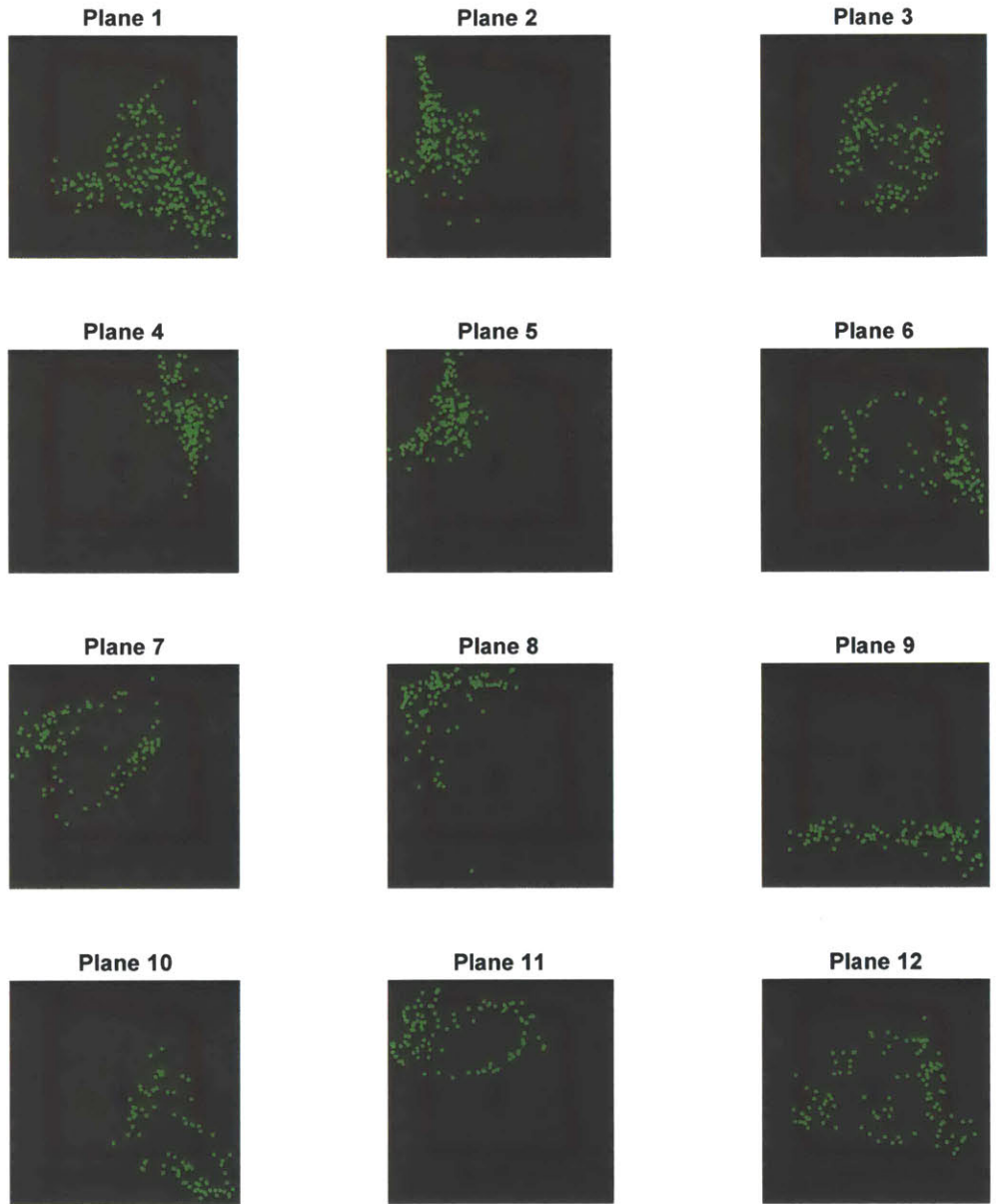
Figure 3-15: Matched keypoints in 12 planes identified using sequential RANSAC superimposed on T. Green markers represent matched keypoints in T.
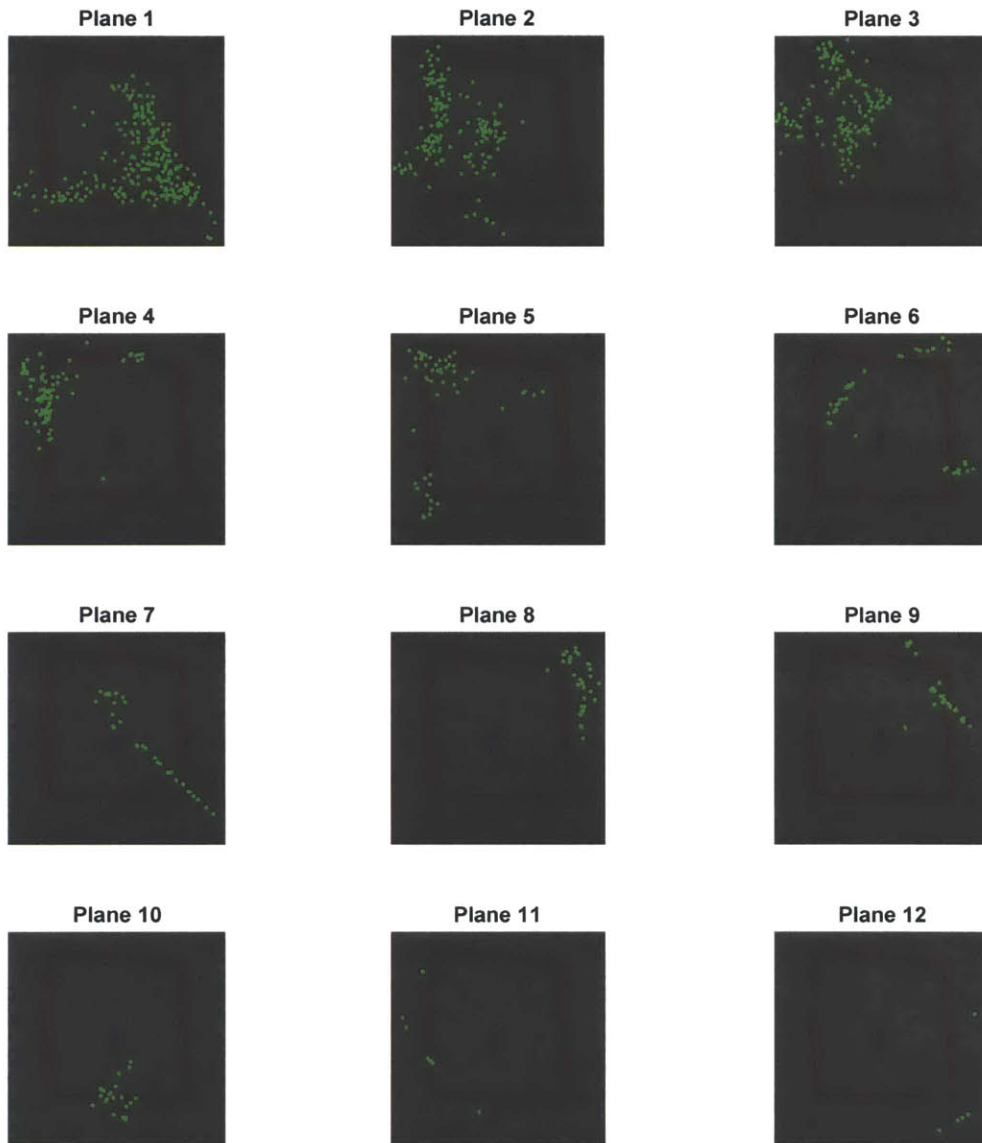
Figure 3-16: Matched keypoints in 12 planes identified using MultiRANSAC super-imposed on T. Green markers represent matched keypoints in T.
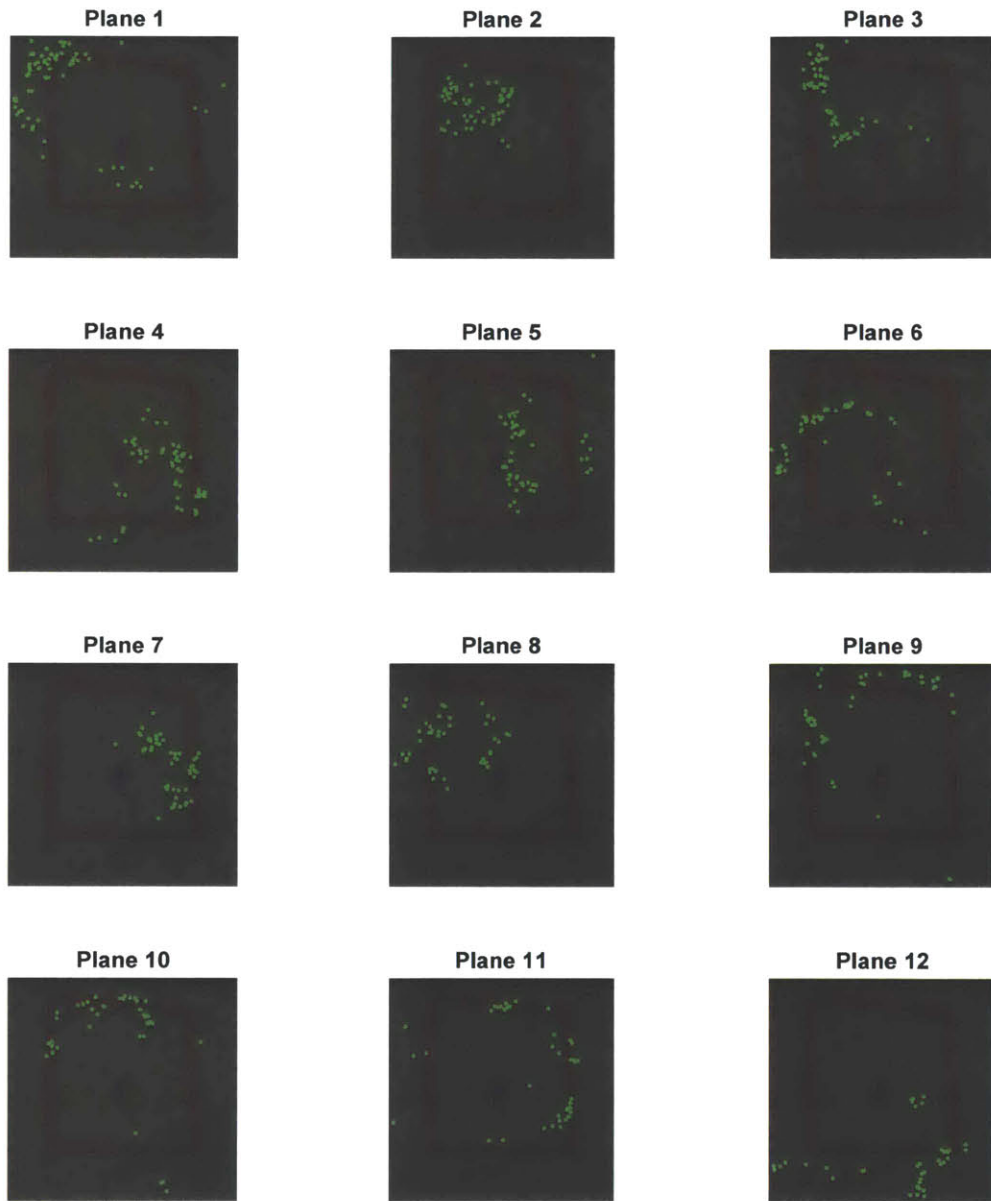
Figure 3-17: Matched keypoints in first 12 planes identified using J-linkage superimposed on T. Green markers represent matched keypoints in T.

skin is that the stretching of the skin must not be too large to cause the microrelief appearance to change too much.

The number of matched keypoints detected in each plane for each MPD method is shown in Figure 3-19. Sequential RANSAC detected 3 planes containing more than 50 matched keypoints each. Although MultiRANSAC was able to detect more keypoints than Sequential RANSAC, it detected many planes with very small amount of keypoints; this increased the chance that any of these keypoints may be in fact an outlier. J-linkage performed poorly in detecting sufficient keypoints in each plane. For instance, its first plane contained only 9 keypoints. Due to low number of initial matched keypoints, the cluster membership threshold η was intentionally set to 6 in order to obtain favorable results. This implies that J-linkage is susceptible to outliers in the case of difficult, noisy images. Hence, by number of keypoints identified, Sequential RANSAC method performed the best in artificially stretched skin image pair.

3.2.5 Real Time-Lapsed Skin Images

Another application of nonrigid skin registration is that of registering the same skin scene imaged at months apart. This allows accurate computer-aided comparison to be made between skin samples. The target image T was taken on 09/29/2015 (Figure 3-20a) and the source image S was taken on 02/23/2016 (Figure 3-20b), after 146 days. We superimposed both images on each other in Figure 3-20c to observe the misalignment. Note that the red square outline on the skin was reapplied on different days, so it was at different locations on the skin. Next, rigid registration was applied. Figure 3-20d shows that 15 matched keypoints were found during rigid registration and the transformed image S_H is shown in Figure 3-20e. Figure 3-20f shows that rigid registration was able to roughly align pairs of images despite their different camera settings. Both Sequential RANSAC and MultiRANSAC were able to identify keypoints all across the ROI in high densities, while J-linkage does not manage to identify keypoints in the top-right corner of the ROI (Figures 3-20g, j and m). Registered images are seen in 3-20h, k and n. Figures 3-20i, l and o display
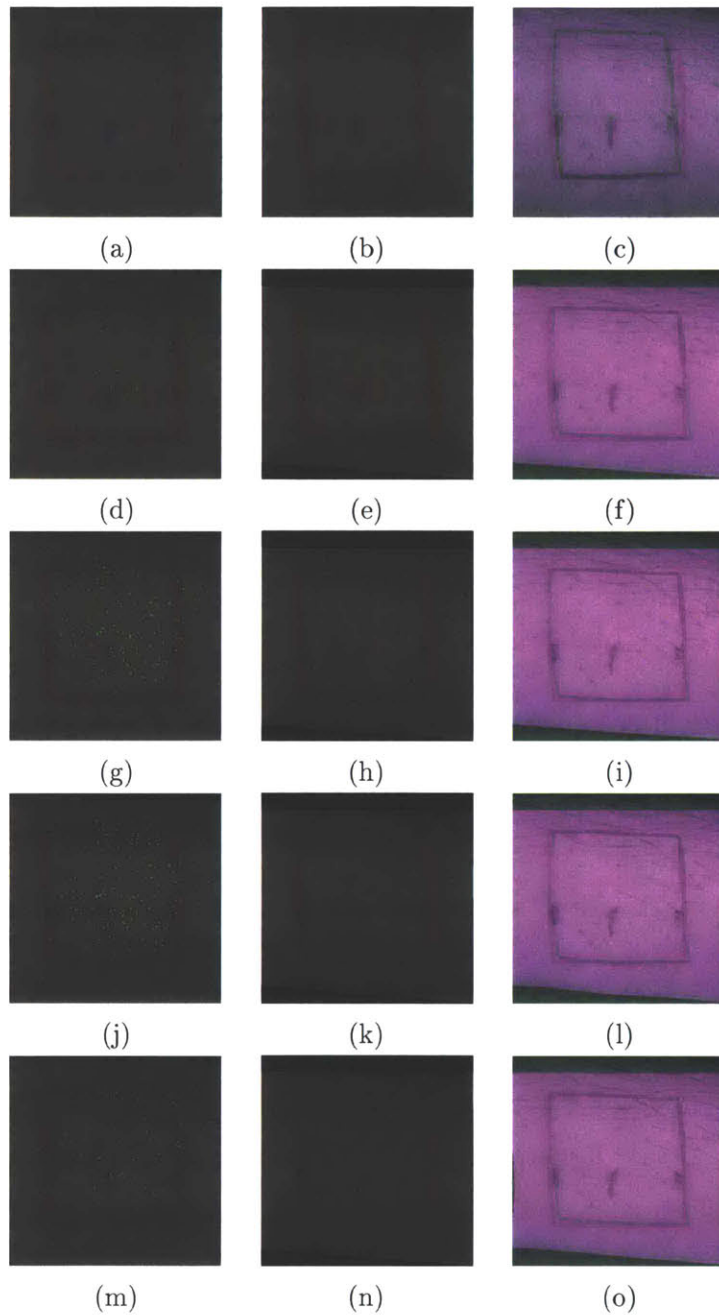
Figure 3-18: Registration result for skin with and without artificially applied stretch. (a-c) Input Images: (a) Target Image T. (b) Source Image S. (c) Superimposed T and S. (d-f) Rigid registration: (d) Matched keypoints superimposed on T. (e) Rigid registration result S_H (f) Superimposed T and S_H. (g-i) Sequential RANSAC: (g) Matched keypoints superimposed on T. (h) Registered image R. (i) Superimposed T and R. (j-l) MultiRANSAC: (j) Matched keypoints superimposed on T. (k) Registered image R. (l) Superimposed T and R. (m-o) J-linkage: (m) Matched keypoints superimposed on T. (n) Registered image R. (o) Superimposed T and R.
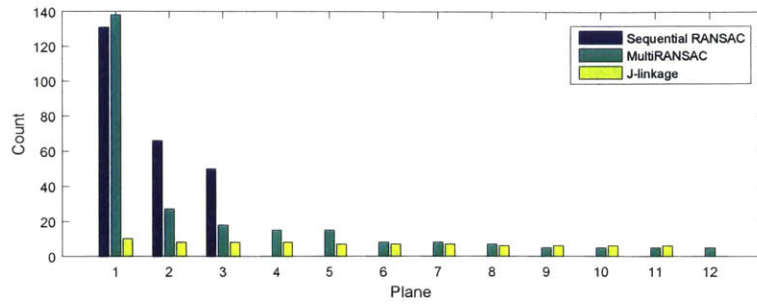
Figure 3-19: Number of keypoints in each plane for skin with and without artificially applied stretch.

the alignment between target and registered images. While the registration result from Sequential RANSAC and J-linkage appeared satisfactory, the result from MultiRANSAC was affected by outlier matched keypoints, as evidenced by some degree of unnatural deformation of the skin on the red square outline.

The number of matched keypoints detected in each plane for each MPD method is shown in Figure 3-21. Sequential RANSAC detected 3 planes containing more than 36 matched keypoints each. Similar to the case of artificially stretched skin, MultiRANSAC detected many planes with very small amount of keypoints. Some of these keypoints were outliers that caused unnatural deformations in the registration result, as we observed in Figure 3-20k. Again, like in the case of stretched skin, J-linkage performed poorly in detecting sufficient keypoints in each plane. The cluster membership threshold η was set to 6 so as to obtain acceptable results.

In general, this example shows that MPD-based registration methods are able to register images of skin reasonably well despite interfering visual markers being present, as the the visual markers were not used in this evaluation. This is similar to the case of registering pigmented skin lesions (PSL) on skin as the inks did not cause erroneous registration results. Since Sequential RANSAC performed best in this experiment in terms of outlier rejection, as future work, we consider testing the Sequential RANSAC-based nonrigid registration algorithm on skin images with PSLs.

76
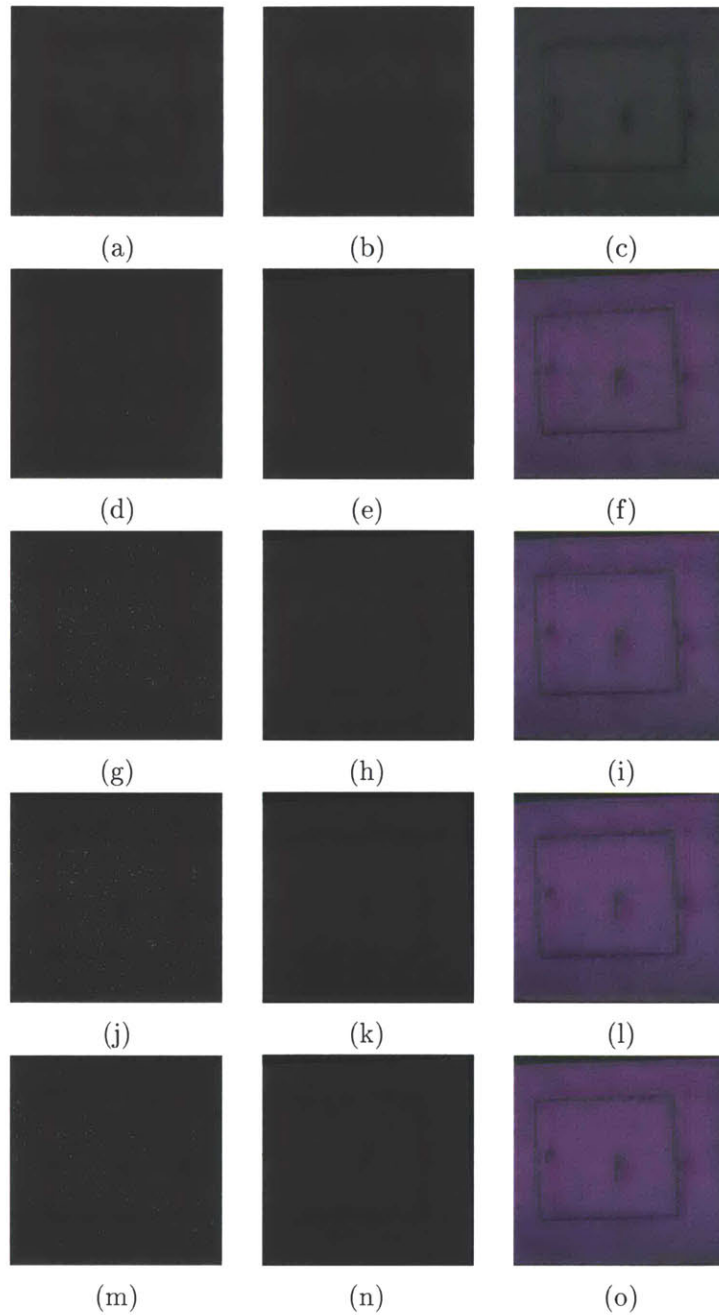
Figure 3-20: Registration result for time-lapse skin images. (a-c) Input Images: (a) Target Image T. (b) Source Image S. (c) Superimposed T and S. (d-f) Rigid registration: (d) Matched keypoints superimposed on T. (e) Rigid registration result S_H (f) Superimposed T and S_H. (g-i) Sequential RANSAC: (g) Matched keypoints superimposed on T. (h) Registered image R. (i) Superimposed T and R. (j-l) MultiRANSAC: (j) Matched keypoints superimposed on T. (k) Registered image R. (l) Superimposed T and R. (m-o) J-linkage: (m) Matched keypoints superimposed on T. (n) Registered image R. (o) Superimposed T and R.
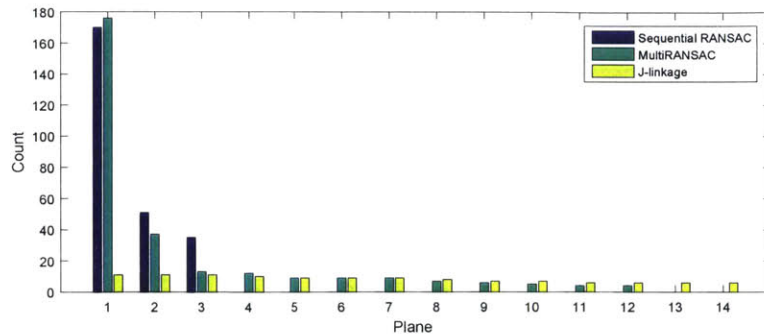
Figure 3-21: Number of keypoints in each plane for time-lapse skin images.

3.2.6 Simulated Images

Given a real base skin image I_{base}, we used the following procedure to obtain an artificially distorted and generated new skin image I_{sim} for subsequent ground truth evaluation. A two-step warping process is used: First, local deformation and global distortion is simulated by a distortion function. Second, a simulated homography is computed based on user-defined camera motion parameters.

We define warping functions that describe the spatial shift of the pixels in I_{base}, the original image, in the x and y coordinate directions:

$$F_x(x) = wf_x(x) \tag{3.19}$$

$$F_y(y) = wf_y(y) \tag{3.20}$$

where w is a scalar warp factor and $f_x(x)$ and $f_y(y)$ are functions that describe the warp in their respective coordinate directions.

Next, to simulate a deliberate stretching effect of the skin, the image is warped with two different scalings in rotated x and y-coordinates about the center coordinates of the distorted image. Specifically, in homogeneous coordinates $\mathbf{x} = [x\ y\ 1]^T$, we

represent this by an affine transformation:

$$\mathbf{H}_{affine}(\mathbf{x}) = \mathbf{T}_{center}\mathbf{R}_s\mathbf{S}\mathbf{R}_z^{-1}\mathbf{T}_{center}^{-1}\mathbf{x} \qquad (3.21)$$

$$\mathbf{T}_{center} = \begin{bmatrix} 1 & 0 & \lfloor N_x/2 \rfloor \\ 0 & 1 & \lfloor N_y/2 \rfloor \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{R}_s = \begin{bmatrix} \cos\theta_s & -\sin\theta_s & 0 \\ \sin\theta_s & \cos\theta_s & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{S} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where the image size is $N_x \times N_y$, θ_s is the rotation angle of the coordinate system and s_x and s_y are the scale factors the in x and y-directions respectively. By assuming that the original camera viewpoint is facing downwards (along the z-axis), perpendicular to the surface at the center coordinates, the 6 degrees-of-freedom of the camera can be set to simulate different viewpoints of the image plane using a homography transformation. This allows us to perturb the viewpoint. The camera's roll θ_r, pitch θ_p and yaw θ_y allows us to construct a rotation matrix $\mathbf{R}_{cam}$:

$$\mathbf{R}_{cam} = \mathbf{R}_x\mathbf{R}_y\mathbf{R}_z \qquad (3.22)$$

$$\mathbf{R}_x(\theta_r) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_r & -s_r \\ 0 & s_r & c_r \end{bmatrix} \qquad \mathbf{R}_y(\theta_p) = \begin{bmatrix} c_p & 0 & s_p \\ 0 & 1 & 0 \\ -s_p & 0 & c_p \end{bmatrix} \qquad \mathbf{R}_z(\theta_y) = \begin{bmatrix} c_y & -s_y & 0 \\ s_y & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$. The camera's translation parameters t_x, t_y and t_z are manually adjusted so as to ensure that the image is approximately centered after camera rotation. Let the focal length be denoted by f_l. The simulated homography by camera motion $\mathbf{H}_{cam}$ is given by:

$$\mathbf{H}_{cam} = \mathbf{T}_{center}\mathbf{H}_{uncentered}\mathbf{T}_{center}^{-1} \qquad (3.23)$$

$$\mathbf{H}_{uncentered} = \begin{bmatrix} f_l r_{1,1} & f_l r_{1,2} & -f_l(r_{1,1}t_x + r_{1,2}t_y + r_{1,3}t_z) \\ f_l r_{2,1} & f_l r_{2,2} & -f_l(r_{2,1}t_x + r_{2,2}t_y + r_{2,3}t_z) \\ r_{3,1} & r_{3,2} & -(r_{3,1}t_x + r_{3,2}t_y + r_{3,3}t_z) \end{bmatrix} \qquad (3.24)$$

where $r_{i,j}$ are entries in $\mathbf{R}_{cam}^T$. The final simulated homography $\mathbf{H}_{sim}$ is:

$$\mathbf{H}_{sim}(\mathbf{x}) = \mathbf{H}_{cam}\mathbf{H}_{affine}\mathbf{x} \qquad (3.25)$$

Finally, Delaunay triangulation [50] is used to interpolate simulated image I_{sim} in RGB.

3.2.7 Accuracy Evaluation Criteria

In the following sections, we used the default parameters for simulated image S3 (resolution 2040×2040) in Table 4.2 for rigid skin registration evaluation in Chapter 4.2.2. The base image I_{base} is shown in Figure 4-5a and distorted simulated image I_{sim} is shown in 4-5d. For an illustration of the combination of planar models found in image pair I_{base} and S3, the matched keypoints found by Sequential RANSAC in their corresponding planes denoted by different colors and markers are marked on image I_{base}. This is shown in Figure 3-22. Alternative planar models leading to different region shapes may be found if another instance of the skin registration algorithm is run.
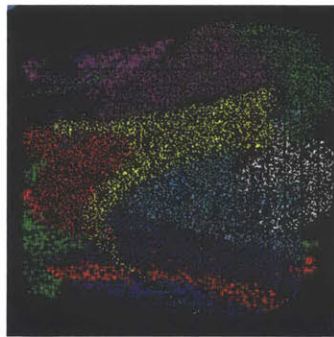
Figure 3-22: Matched keypoints in corresponding planes for simulated image S3. Matched keypoints belonging different planes may be differentiated by their color and marker shape. The matched keypoints are superimposed on base image I_{base}.

Accuracy of registration is evaluated based on the root-mean-square error (RMSE) of the dense set of corresponding microrelief junction points J_b contained within a user-defined ROI in the domain of I_{base}, as displayed in Figure 3-23b. Microrelief

80

junction point extraction is discussed in Chapters 4.1.4 and 4.1.5. The known transformation of S3 was applied to 22915 junction points in a square ROI region in I_{base} in domain range $x = [451, 1150]$ and $y = [451, 1150]$. We denote the set of transformed junction points as J_t. The junction points correlate with actual physical features on the skin surface, thus they are meaningful points for registration evaluation. In order evaluate the error, J_t is back-transformed by registration into the set of back-transformed junction points J_t'. Denoting the simulated transformation as $\mathbf{T}_{sim}$, the RMSE in pixels is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{|J_b|} \sum_{\mathbf{x} \in J_b} \|\mathbf{T}(\mathbf{T}_{sim}(\mathbf{x})) - \mathbf{x}\|^2} \tag{3.26}$$

A lower RMSE indicates better accuracy of registration.

For simulated images, many initial matched keypoints could be identified by grid-based NNDR as image intensities were not corrupted by noise, varying illumination levels or different camera settings. Thus, we set a high $\eta = 40$ for J-linkage to remove many small clusters in the following sections.

3.2.8 Varying the Number of Levels in Multilevel B-Spline

The multilevel B-spline algorithm [52] approximates the deformation field $\mathbf{T}_{local}(\mathbf{x})$ by reducing the Euclidean distance error between paired keypoints as the level l increases. For an image of resolution 2040×2040, there are a maximum of $L = 10$ levels. The resolution of the B-spline grid is approximately doubled every level, thus smaller deformations are modeled as l increases. High resolution grids are sensitive to the effect of outlier matched keypoints, which would decrease registration accuracy. On the other hand, low resolution grids are not sensitive to small deformations; this also decreases the accuracy of registration. Thus, each level of multilevel B-spline introduces different degrees of spatial regularization that helps reduce the effect of outliers in the data. Consequently, it is desirable to find the optimal maximum level L.

For this evaluation, we constructed sets of keypoints K_o with a desired percentage

p_o of outliers. First, we used Sequential RANSAC to find the matched keypoints in image pair I_{base} and S3. A random sample of 10% of all 10322 matched keypoints K_f was taken so as to construct a set of sparsely distributed keypoints K_{sp}. Second, we added outlier keypoints to K_{sp} to make up $p_o\%$ of K_o. Outlier keypoints were taken from a random sample of the remaining paired keypoints in K_f. The keypoints in the domain of I_{base} were randomly perturbed up to 5 pixels in L_1 distance. We illustrate K_o with $p_o = 10$ in Figure 3-23a. Third, we ran multilevel B-spline approximation with values of $p_o = \{0, 5, 10, 15, 20\}$ for K_o and maximum levels $L = \{1, \ldots, 10\}$. Fourth, the RMSE of the junction points were evaluated using Equation 3.26. The results are plotted in Figure 3-24a. The minimum RMSE can be found at $L = 3$ for almost all p_o. Hence, we set $L = 3$ for multilevel B-spline for the remaining evaluation in this chapter. The plot also confirms that the smaller deformations at higher levels are sensitive to outlier matched keypoints. As expected, a higher p_o led to an overall increase in RMSE. Also, when we observe the residuals or mean Euclidean distance between matched keypoints after registration in Figure 3-24b, their values decreased as l increased. Hence, at high L, the effect of outliers on the RMSE outweighs the effect of minimizing distance between paired keypoints.

(a)   (b)
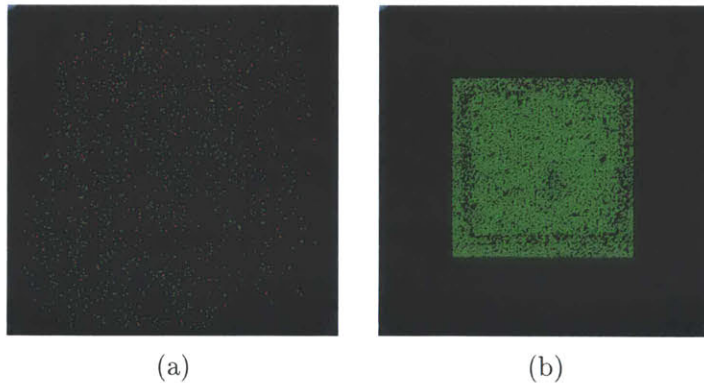
Figure 3-23: (a) Matched keypoint set K_o with 10% outlier keypoints. Green points represent inliers and red points represent outliers. (b) Dense set of junction points within the ROI in the domain of I_{base}.
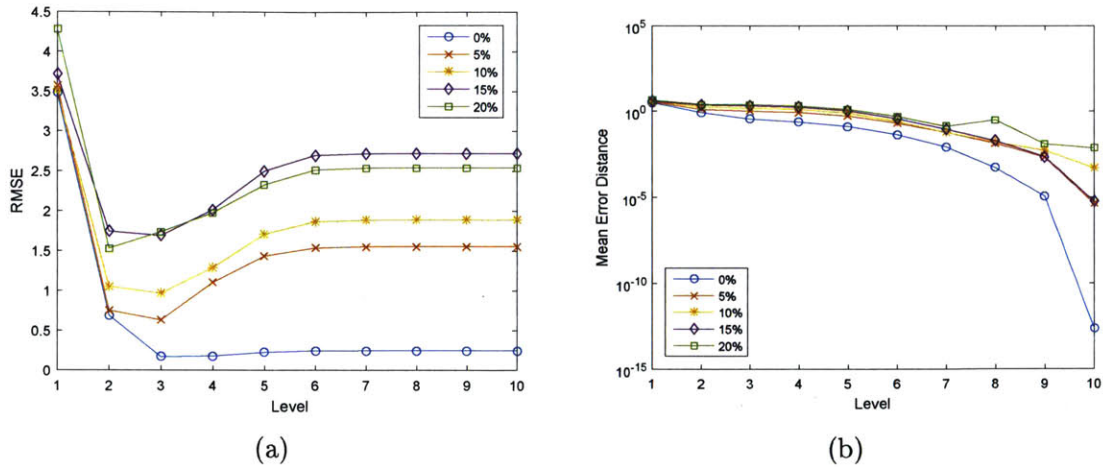
Figure 3-24: RMSE of junction points and mean distance error of matched keypoints for varying maximum B-Spline grid levels L and outlier percentage p_o. (a) RMSE of junction points. (b) Mean distance error of matched keypoints.

3.2.9 Varying Rotation and Distortion

Next, we evaluated skin registration for robustness to various in-plane rotations $\theta_y = \{-10°, -5°, 0°, 5°, 10°\}$ and various warp factors $w = \{2, 2.5, 3, 3.5, 4, 4.5, 5\}$. The RMSE for Sequential RANSAC, MultiRANSAC and J-linkage are plotted in Figures 3-25a, 3-25b and 3-25c respectively. In general, the results show that in-plane rotations do not affect registration accuracy as distortion due to rotation being mostly accounted for by rigid registration. Additionally, RMSE increased as distortion scale is increased. Sequential RANSAC maintained low levels of RMSE below 0.3 for all distortion scales, while MultiRANSAC performed as well as Sequential RANSAC at low distortion scales but was more affected by increasing levels of distortion. However, the accuracy of J-linkage was comparatively poorer than both Sequential RANSAC and MultiRANSAC at all warp factors. To summarize, the mean and median RMSE of each method at all rotations and distortion scales are given in Table 3.2. Overall, it is clear that Sequential RANSAC has the best accuracy and exhibited the least variability of error.
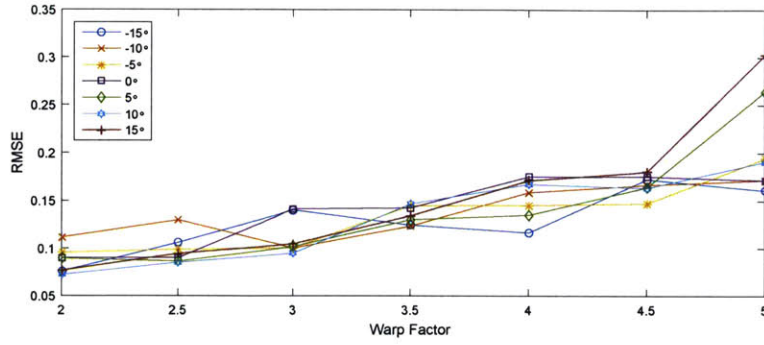
(a)
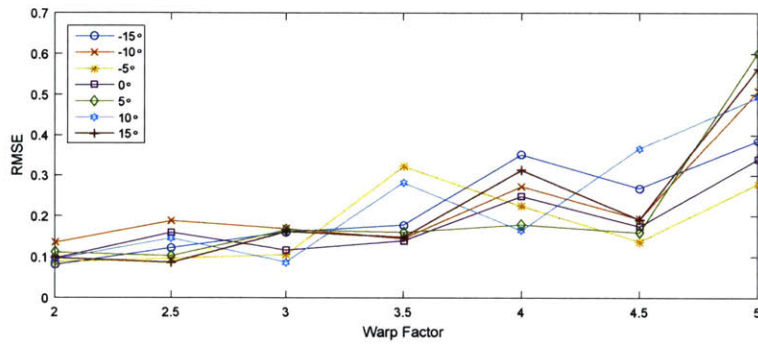
(b)
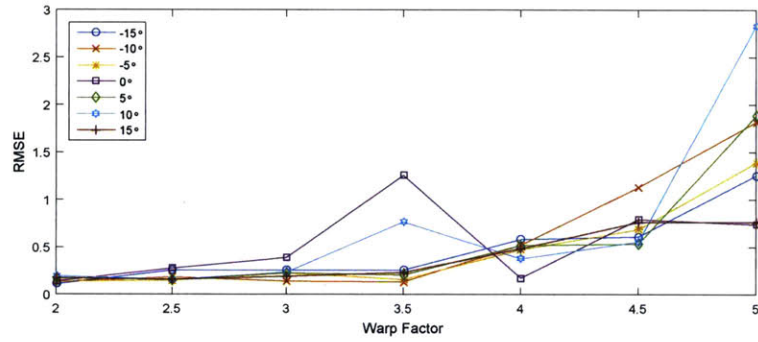
(c)

Figure 3-25: RMSE against warping factor and in-plane rotation for skin registration. (a) Sequential RANSAC. (b) MultiRANSAC. (c) J-linkage.

Table 3.2: Mean and median RMSE across all in-plane rotations and distortion scales.

RMSE	Sequential RANSAC	MultiRANSAC	J-linkage
Mean	0.1374	0.2117	0.5298
Median	0.1351	0.1647	0.2579

3.2.10 Varying Noise Levels

In this test, we quantified the accuracy of non rigid registration with respect to the amount of noise in skin images. Assuming that simulated image S3 was not corrupted by noise initially, we added white Gaussian noise to rigidly registered image S3 and denote this image as $S3_H$. Then, a fixed amount of salt and pepper noise is added. There is a $p_{sp} = 0.05$ probability of any pixel in S3 having an amount $t_{sp} = 3$ added to or subtracted from the local mean intensity in each color channel (intensity values are integers in range $[0, 255]$). Half of the chosen pixels have t_{sp} added to their values and the other half have t_{sp} subtracted from their values. Taking additive white Gaussian noise and salt and pepper noise in combination, we simulated sensor noise caused by poor illumination. As a result, fewer SIFT features were detected and consequently, fewer numbers of initial matched keypoints were identified. Examples of noise-corrupted images are shown in Figure 3-26 for various signal-to-noise ratios (SNR).

(a)   (b)   (c)
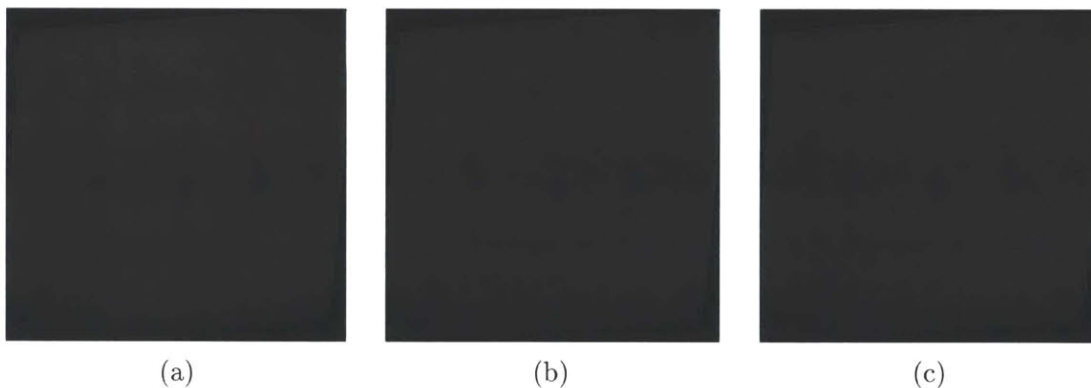
Figure 3-26: Rigidly registered simulated image S3 corrupted by additive white Gaussian noise and salt and pepper noise with various SNRs. (a) SNR=20 dB. (b) SNR=10 dB. (c) SNR=2 dB.

For each MPD method, the RMSE of the registered images with respect to different Gaussian noise levels in dB are plotted in Figure 3-27a. The plot does not reveal the proportional relationship between noise level and number of initial matched keypoints, so we replaced the noise levels with number of initial matched keypoints and plotted the results in Figure 3-27b. The RMSE for Sequential RANSAC was consis-

85

tently lower than those of other methods for above 1000 initial matched keypoints. MultiRANSAC exhibited RMSE values much higher than that RMSE of Sequential RANSAC or J-linkage for 2000 matched keypoints and above; the parallel assignment of disjoint sets of keypoints contributed to high numbers of outlier keypoints for MultiRANSAC. The performance of J-linkage excelled when the number of keypoints was high. On the other hand, its usage of clustering the preference set for evaluating models is extremely tricky for low numbers of initial matched keypoints and this gave rise to high amount of outliers at 1000 and below keypoints. To enable J-linkage to produce reasonable results at low numbers of keypoints, we adjusted the cluster membership threshold η to 10 for noise levels of 8 dB and below (equvialently, about 2000 keypoints and below). For all other Gaussian noise levels, η remained at the default value of 40.

3.2.11   Computation Time for Keypoint Matching

Keeping the number of MSS's constant for all 3 MPD methods with the default parameters, we investigated computation time with respect to number of initial matched keypoints. Note that the number of initial matched keypoints is high if two images appear similar at the scale of the microrelief patterns, and vice versa. Using the setup covered in Chapter 3.2.10, the runtime in seconds is plotted against the number of initial matched keypoints for the 3 MPD methods (Figure 3-28). The runtime of both Sequential RANSAC and MultiRANSAC showed a linear relationship with the number of initial matched keypoints. This is because RANSAC only requires a single pass through all initial matched keypoints when it computes the CS. Also, Sequential RANSAC has fewer computational overheads per pass and so performed better than MultiRANSAC overall. However, the runtime of J-linkage has quadratic time complexity with respect to number of initial matched keypoints because of agglomerative clustering and this is reflected in the plot in Figure 3-28. Thus, we expect a pair of images with similar illumination levels and minimal distortion to take the longest time to match, while severely degraded image pairs take the shortest time to match.
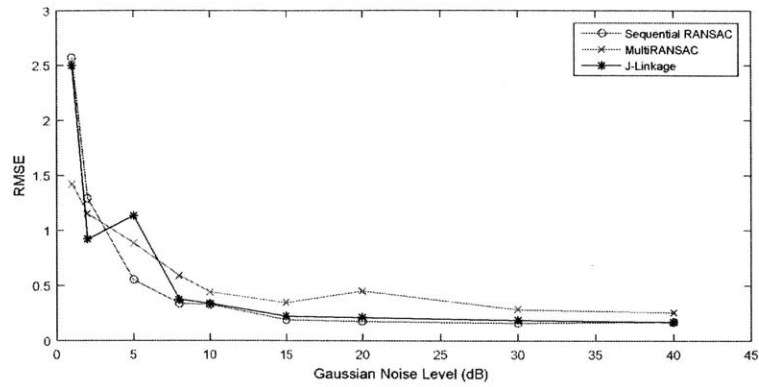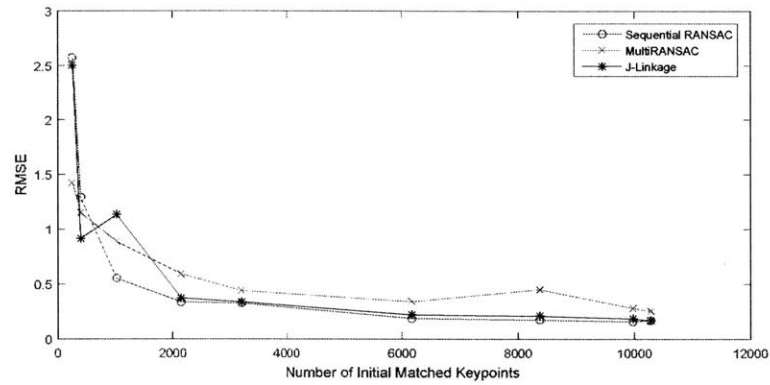
Figure 3-27: (a) RMSE against Gaussian noise levels for different MPD methods. (b) RMSE against number of initial matched keypoints for different MPD methods.
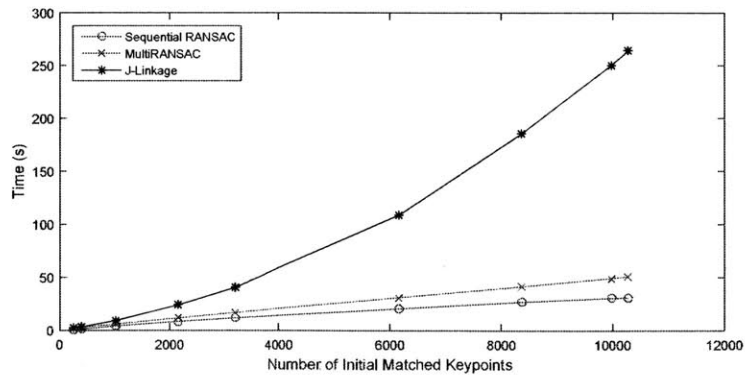
Figure 3-28: Runtime against number of initial matched keypoints for 3 MPD methods.

3.3 Final Design and Discussion

The final design for skin registration uses Sequential RANSAC for MPD keypoint detection. It performed strongly in our accuracy and runtime experiments. It also gave the most visually pleasing registration results and relatively high numbers of final matched keypoints for the differing viewpoints, artificially stretched skin and time-lapsed real image pairs. It has the additional advantage of being simple to implement and having few parameters that do not need to be manually adjusted in many cases.

MultiRANSAC's concept of the parallel assignment of disjoint sets of keypoints and the need to set a prior number of planar models did not work well in general. This is because the number of ground truth planes is unknown in a real skin scene. As a result, MultiRANSAC consistently extracted too many planes with few matched keypoints and consequently, outliers were consistently extracted. Hence, its accuracy was generally worse than that of Sequential RANSAC. Also, it performed slightly worse than Sequential RANSAC in runtime.

J-linkage performed well for a large set of initial matched keypoints and is easy to implement. However, when few initial matched keypoints were present in pairs of noisy skin images, the cluster membership threshold η required manual intervention. Furthermore, agglomerative clustering requires quadratic time complexity with respect to number of initial matched keypoints, so it is slow compared Sequential RANSAC for large numbers.

We have shown that our Sequential RANSAC-based skin registration method achieves highly accurate registration of skin images using only skin features as landmarks. It is robust to in-plane rotations and realistic local and global distortion scales, and fairly invariant to image noise and camera and lighting settings. Its local motion model, Multilevel B-spline, is also able to reduce the effect of outlier landmarks identified by Sequential RANSAC in order to ensure more accurate final alignment.

In this chapter, we did not consider how hair-occluded skin affects registration results. However, in Chapter 6, we demonstrate that by preprocessing skin images

with an digital hair removal (DHR) algorithm and inpainting the skin image, we are able to obtain fairly accurate registration with our nonrigid registration algorithm despite the presence of thick and dense hairs.

In typical real skin image pairs, nonrigid registration is fast and reliable; it only takes about 10 seconds to run and rarely fails (i.e. the registered image R appears unnaturally warped at the global scale). However, global registration remains slow, requiring around 5 seconds to 15 seconds, depending on the number of initial matched keypoints present. Most of the increased runtime can be attributed to traditional NNDR (see Table 3.1). To improve runtime, alternative global registration methods need to be explored. Furthermore, the current rigid registration method is susceptible to registration failure on noisy image pairs. Therefore, future work includes developing a global registration method for more consistent successful results.

Our implementation for Sequential RANSAC did not remove isolated matched keypoints in each model. These are keypoints in some planes that have few neighboring keypoints in its locality that are more likely to be outliers than keypoints located in densely populated local areas. This improvement are trivial to implement but they are could further improve the accuracy of skin registration.

In our evaluation, we have only considered 3 MPD methods for keypoint matching. A number of alternative MPD methods can be found in the literature. As future work, we consider evaluating other MPD methods such as histogram of residuals [107], Mercer kernel fitting [14] and graph cuts [36] approaches.

Chapter 4

Microrelief Junction Point Matching

As an application to rigid skin registration, we present a concentric ring-based non-rigid point pattern matching (PPM) method for microrelief junction point matching (MJPM). The MJPM procedure was developed as a first step towards eventually developing a framework for quantitative assessment of the stability and evolution of the microrelief of skin over time. This work was developed in close collaboration with Dr. Xian Du. In this chapter, we describe the algorithm in detail and evaluate results from our proof-of-concept.

4.1 Algorithm Design

4.1.1 Overview

Figure 4-1 illustrates the flowchart of our MJPM procedure that takes a source image S and target image T as inputs and outputs the set of matched junction points. It comprises three major sequential steps: homography and keypoint estimation, microrelief junction point detection and PPM. First, the images are registered by feature-based homography transformation, during which the number of matched keypoints found is augmented using a two-step process. After rigid registration, the transformed source image S_{H_2} and a set of matched keypoints found in T and S_{H_2} are obtained. Since T and S_{H_2} are well-aligned in a small, locally planar region, we use

the convex hull of the matched keypoints in these images to designate our region of interest (ROI) for PPM. Second, the microrelief structure of T and S_{H_2} are extracted. Next, junction points are detected from the microrelief structure based on crossing numbers. Third, the ROI is split into multiple concentric ring layers, each conforming to the shape of the convex hull. For each layer, an assignment problem between junction points involving cyclic order preserving constraints and shape context descriptor cost is solved by dynamic programming.

4.1.2 Homography and Keypoint Matching

Because the ROI for PPM is constructed based on the convex hull of matched keypoints between registered source image S_{H_2} and target image T, we require a good set of matched keypoints: the set should form a single, spherical cluster with uniform distribution of points in order to ensure good alignment of points enclosed by the ROI. Registration is achieved by a planar homography transformation, which models viewpoint changes of the skin scene or small stretches of the skin surface. For the rest of this section, we explain how the rigid registration method in Chapter 3 is modified to find a good set of matched coplanar keypoints in the scene and its homography estimate.

First, taking source image S and target image T as inputs, we use the same keypoint detection, nearest neighbor ratio test (NNDR) test, random sample consensus (RANSAC) and nonlinear iterative refinement procedure presented in Chapters 3.1.2 and 3.1.3. The process gives us the first homography estimate $\hat{H}_1$ and its corresponding globally registered image S_{H_1}. However, the matched keypoints identified at this stage may not be sufficient in number or evenly distributed to ensure accurate alignment of all points within the proposed ROI. Therefore, the number of matched keypoints needs to be augmented.

A second round of keypoint detection and matching, this time between images S_{H_1} and T, is carried out. After extracting Scale Invariant Feature Transform (SIFT) features from S_{H_1}, SIFT features from S_{H_1} and T are matched using the fast grid-based initial keypoint matching method described in Chapter 3.1.4. As we concluded
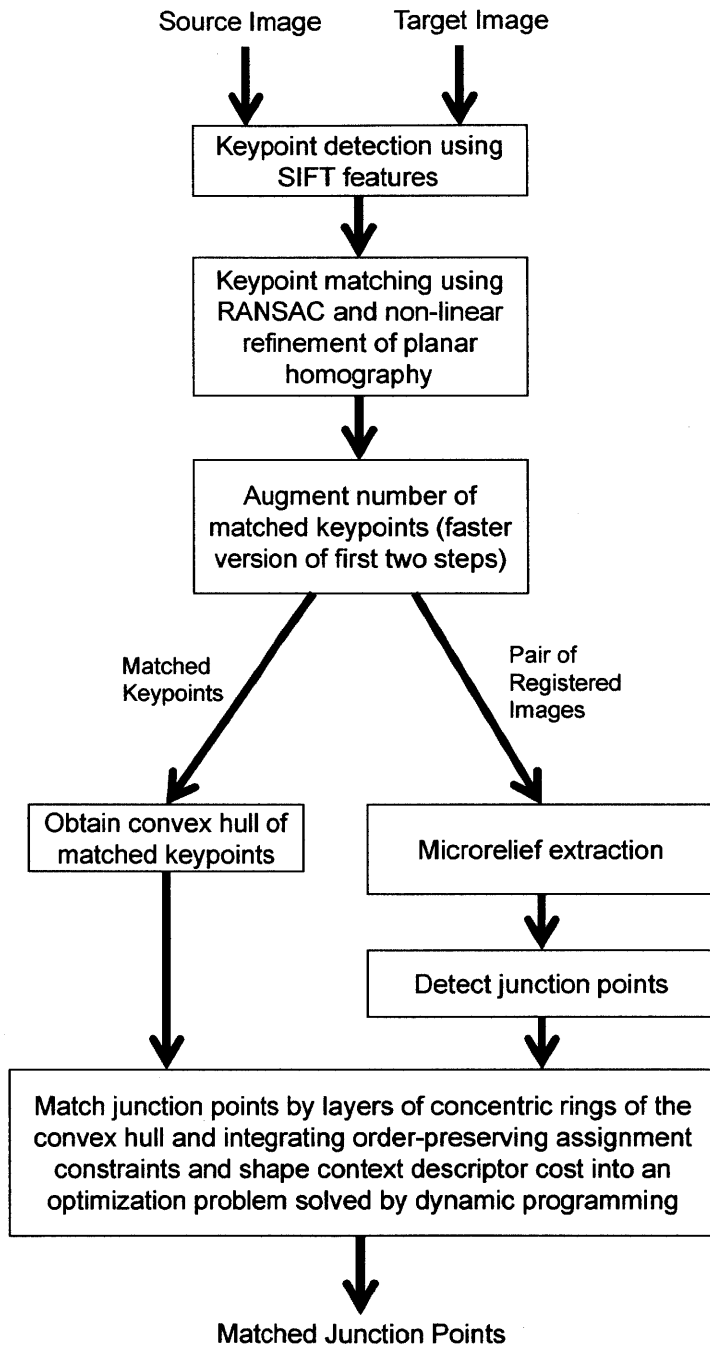
Figure 4-1: Algorithm flowchart for microrelief junction point matching procedure.

previously, the grid-based matching method greatly increases the number of initial matched keypoints in comparison to the first round of keypoint detection and matching. It also helps that SIFT keypoints not previously matched are now more likely to be matched due to the homography correction applied. Again, through RANSAC and iterative refinement of the homography estimate, we obtain a list of inlier matched keypoints and a final homography estimate $\hat{H} = \hat{H}_2 \hat{H}_1$, where $\hat{H}_2$ is the homography estimate from the second round. Next, using Equation 3.2 and $\hat{H}_2$ as the homography, pixel locations in S_{H_1} are warped to obtain pixel locations in the final globally registered image S_{H_2}. The RGB pixel intensities in S_{H_2} are calculated by bicubic interpolation.

Equation 3.2 is also used to transform matched keypoints in the domain of S_{H_1} to the domain of S_{H_2}. In many cases, keypoints matches identified are clustered in the ROI, but sometimes, small coplanar regions in the scene may also cause smaller clusters to exist elsewhere. Since a single cluster of evenly distributed matches is required for the next stage, we remove a keypoint unless there are at least K_n keypoints found within a radius K_r from that keypoint. Large values of K_n and small values of K_r tend to filter out keypoints in sparsely populated regions of keypoints, while small values of K_n and large values of K_r tend to retain more isolated keypoints. As such, the final set of matched keypoints in the domains of T and S_{H_2} is obtained.

4.1.3 Convex Hull Registration

The matched keypoints in image pair T and S_{H_2} approximately enclose a locally planar region in 2D Euclidean space with good alignment between both images. The convex hull, defined as the smallest convex set in 2D space that contains the matched keypoints is used to approximate the locally planar region. Geometrically, the Delaunay triangulation [50] of a set of points in 2D space results in a set of triangle edges connected like a network, with the points as nodes. The edges that are shared only by a single triangle in the network reveal the convex hull. As global registered image S_{H_2} is an estimate of target image T, a globally registered ROI in T and S_{H_2} is acquired.

94

4.1.4 Microrelief Structure Extraction

The purpose of this stage is to extract the valley-like structures of the microrelief from noisy skin images T and S_{H_2}. The following operations are applied on both images. First, the RGB image are converted to YUV (BT.709) and the intensity component Y is extracted. Second, background normalization is applied. A 13×13 box linear filter is applied iteratively for 30 times to approximate a Gaussian blur. The resulting blurred image is subtracted from the original image. Third, the bottom of flat valleys and top of flat mesas in the image are enhanced. The local standard deviation is calculated for each pixel, using a 5×5 window size. Then, the gradient magnitude (using central difference) and discrete Laplacian images are computed from the local standard deviation image. Fourth, a global threshold is applied to the gradient magnitude image to find pixels that have small gradient magnitude. The threshold used is the mean of the gradient magnitude image. Next, since we are looking for minima on the local standard deviation surface, pixels in the binary image where the discrete Laplacian is greater than 0 are retained. Thus, this step extracts the bottom of flat valleys. Fifth, morphological thinning is applied to the binary image to reduce the valley structures into one pixel width structures. Subsequently, isolated pixels are removed and holes in the skeleton are filled. Sixth, the microrelief structure is extracted by identifying and retaining the largest connected component in the skeleton image.

4.1.5 Junction Point Detection

The microrelief structure is a network that is connected by junction points, which by our postulate in Chapter 2.2.3, forms a set of skin features that are stable over a time scale of a few years. Hence, matching skin patches is akin to matching two networks and their nodes. In the same vein as minutiae extraction in fingerprint analysis [108], the Rutovitz crossing number (CN) [81] is used to extract the junction points. At each pixel of the skeleton image, the eight neighbors of the pixel are scanned in an

anticlockwise manner. The CN for a pixel P is given by

$$\text{CN} = \frac{1}{2} \sum_{i=1}^{8} |P_i - P_{i+1}| \qquad (4.1)$$

where P_i represents the binary pixel value in the neighborhood of P and $P_1 = P_9$. For illustration, the neighboring pixels are labeled as shown in Figure 4-2a. In Figure 4-2b, we see that various CN values denote various properties. We define junctions points to be bifurcation and crossing points. Thus, if CN $\geq$ 6, pixels in the microrelief structure are retained and labeled as junction points.
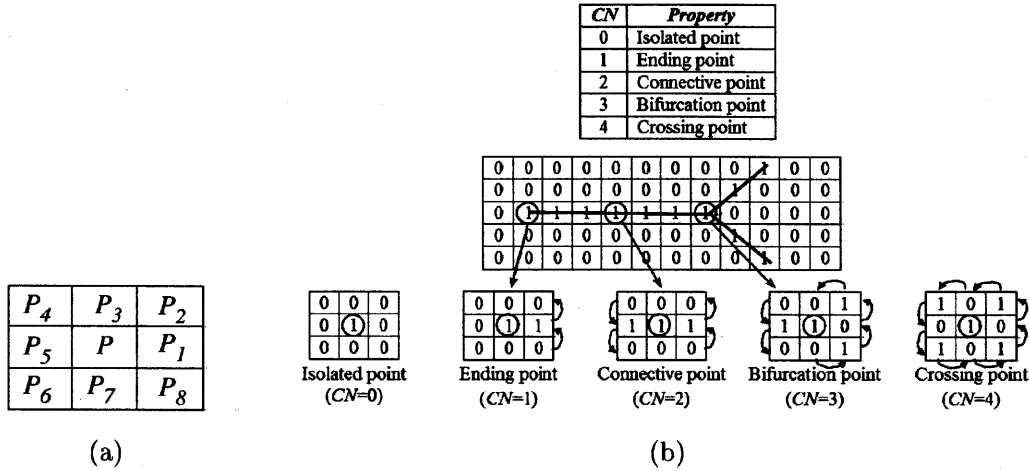
Figure 4-2: Illustration of the Rutovitz Crossing Number (CN) and properties. (a) Labels of binary pixel values P_i in neighborhood of pixel P. (b) CN and its associated properties of skeleton image at pixel P. Images adapted from [108].

4.1.6 Point Pattern Matching

Rather than matching all the junction points within the ROI in T and S_{H_2} directly, we match the junction points found separately in multiple concentric ring layers with fixed width thickness r_w, each ring conforming to the shape of the convex hull. We list some reasons for using this approach: (1) Directly matching two large sets of junction points involves high computational cost and memory cost; the problem is made more manageable by matching junction points found in ring layers. (2) The skin surface can

be assumed to be homogeneous in proper stretching, so sets of microrelief junction points in two rings to be matched are assumed to contain mostly the same junction points. (3) The junction points in each ring can be ordered in a circular manner so as to define single, closed contour. This allows for the cyclic order preserving constraint to be implemented. The cyclic order preserving constraint increases the dissimilarity between junction points that do not obey cyclic ordering; this increases the accuracy and geometric fidelity of junction point matching.

Cyclic Order Preserving Assignment Problem

A modification of the standard assignment problem is made to require assigned junction points to conform to the cyclic ordering inherited from the closed contour. This modification is known as the cyclic order preserving assignment problem (COPAP), introduced by Scott and Nowak in [84], in which we describe below for our context.

Each concentric ring in the ROI requires solving a COPAP. In the following description, we describe the mathematical formulation. Let $\{p_i \mid p_i \in \mathbb{R}^2, i = 1, \ldots, m\}$ and $\{q_j \mid q_j \in \mathbb{R}^2, j = 1, \ldots, n\}$ denote junction point locations found in images S_{H_2} and T respectively. Without loss of generality, assume that $m \leq n$. The dissimilarity between points p_i and in q_j can be measured by a non-negative cost c_{ij}. Then, a cost matrix $\mathbf{C} = (c_{ij})_{m \times n}$ denotes all possible pairing costs between the two junction point sets. To deal with the missing points and outliers in the junction point sets, nonexistent *dummy points* are introduced. This allows some points to be matched to dummy points. Formally, a *matching* of junction points is a function $\pi : \{1, \ldots, m\} \rightarrow \{0, 1, \ldots, n\}$ such that no value in $\{1, \ldots, n\}$ is assumed more than once. The notation $\pi(i) = 0$ specifies that p_i is not matched with any q_j. The dummy pairing cost is user-specified, denoted by ϵ; it is a constant generally set larger than the maximum value of the non-dummy pairing costs in $\mathbf{C}$. Consequently, the $m \times n$ cost matrix can be extended to an $N \times N$ matrix, where $N = m + n - L$, by padding it with ϵ to effectively add dummy points. Here $L \in [0, m]$ denotes the minimum number of non-dummy pairings in Π_L, the set of all matchings of size at least L. Since the junction point sets are in a closed contour, order them in a clockwise manner

97

respectively. The set of all matchings in Π_L that preserve the order is indicated by Π_L^{cyc}. The COPAP is defined as

$$\min_{\pi \in \Pi_L^{cyc}} \sum_{i=1}^{m} c_{i,\pi(i)} \qquad (4.2)$$

where $c_{i,0} = \epsilon$ for all i.

We give a brief description for solving the COPAP, of which a detailed exposition is found in [84]. The COPAP is solved by repeated solving of a problem known as the linear order preserving assignment problem (LOPAP). The optimal matching is found by solving for m circular shifts of the junction points p_i and taking the solution with lowest cost. We chose the fast $O(mn)$ dynamic programming algorithm presented by [84] to solve the LOPAP when $L = 0$, indicating that there is no minimum matching constraint. This method is advantageous in terms of speed for solving large-scale problems involving point sets with large numbers such as junction point matching. Each LOPAP is solved by dynamic programming using the recursive relation to traverse an $(m + 1) \times (n + 1)$ directed graph:

$$w_{i,j} = \min\{w_{i-1,j} + \epsilon, w_{i,j-1}, w_{i-1,j-1} + c_{i,j}\} \qquad (4.3)$$

$$w_{0,0} = 0 \qquad (4.4)$$

There is no need is determine the cost for all m circular shifts of p_i due to the use of a bisection strategy to solve LOPAP at all shifts. Hence, the COPAP is solved in $O(mn \log m)$ time.

Shape Context Descriptor

Entries of the cost matrix $c_{i,j}$ indicate the amount of dissimilarity between junction points on different sets p_i and q_j. For every junction point, its shape context descriptor [6] is the log-polar histogram of coordinates of the remaining points in the same set. The K-bin normalized histograms are represented by $h_i(k)$ and $h_j(k)$ for points p_i

and q_j respectively. The distance between both points is given by the χ^2 test statistic

$$c_{i,j} = \frac{1}{2} \sum_{k=1}^{K} \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \tag{4.5}$$

4.2 Experimental Results

4.2.1 Implementation and Parameters

The algorithm was implemented in MATLAB Release 2015b on a system with an Intel Xeon E3-1225 v3 3.2 GHz processor. In following sections, we present and discuss our preliminary results for rigid skin registration and MJPM. We also describe our procedure our creating simulated images of distorted skin images for quantitative evaluation.

The RANSAC inlier threshold ζ is set to 1 for both rounds of keypoint matching. If the homography model is a ground truth model, this ensures that registration errors in the local planar regions of the scene are sub-pixel. Next, the number of RANSAC iterations or trials M is set to 1000. This allows for a reasonably high probability of finding the ground truth homography model. The width of the concentric ring layers r_w is set to 20. Setting r_w too small introduces greater matching error if there were large local distortions while having r_w set too large increases the computational time as more combinations of matchings need to be considered. To ensure that all keypoints belong to a high density region of keypoints, the final keypoint refinement parameters are chosen as: $K_n = 20$ keypoints and $K_r = 80$ pixels.

4.2.2 Rigid Skin Registration

Evaluation with Real Images

We demonstrate rigid skin registration with a pair of real time-lapsed skin images. An image pair from the ventral forearm of a male, 23-year-old subject of South Asian descent was used. The target image T was taken on 09/29/2015 (Figure 4-3a) and

the source image S was taken on 02/23/2016 (Figure 4-3b), which is a duration of 146 days or approximately 5 months apart. The RGB images are of resolution 1340×1340.

(a)   (b)
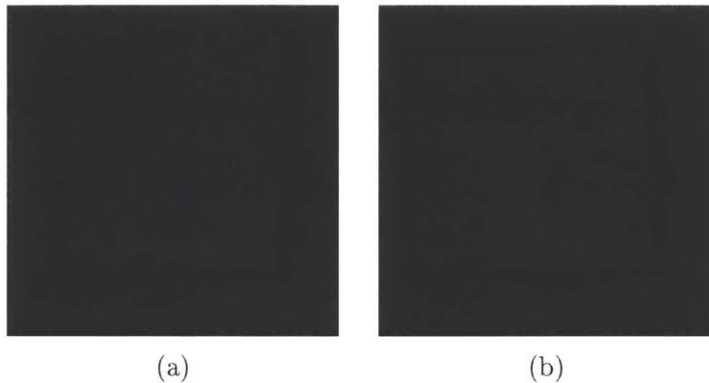
Figure 4-3: Time-lapsed image pair. (a) 09/29/2015 (Day 1). (b) 02/23/2016 (Day 147).

In Figure 4-4, we show the results of matched keypoints after the 1st and 2nd rounds of keypoint matching. Keypoints from the target image T are displayed in red and keypoints from the rigidly registered source images S_{H_1} and S_{H_2} are displayed in blue. The keypoints are superimposed on the images to denote their location. The green image is T and the red images are the rigidly registered source images S_{H_1} and S_{H_2} from both rounds of keypoint matching. The keypoints are not ground truth keypoints as they are extracted from separate images using SIFT. But if we assume that they represent same locations, then it is visually observed that there is only very small sub-pixel error between the matched keypoints. Furthermore, 23 matched keypoints were detected after the 1st round and 79 matched keypoints were detected after the 2nd round. Although we may proceed to the microrelief matching stage at the 1st round with few matched keypoints, a greater quantity of keypoints identified would give us better confidence in accurate convex hull registration. Also, the 2nd round of keypoint matching results in a vast improvement in the number of keypoints; an approximate increase of 4 times the number of keypoints from the 1st round. We mainly attribute this improvement to the grid-based initial keypoint method described in Chapter 3.1.4 used in the 2nd round of keypoint matching as the method implicitly decreases the NNDR threshold. Lastly, the overlapping images

suggest that there is good image alignment in the local regions approximated by the keypoints. Therefore, for this image pair, a visually accurate agreement of keypoints and rigid registration is obtained.

<p align="center">(a)   (b)</p>
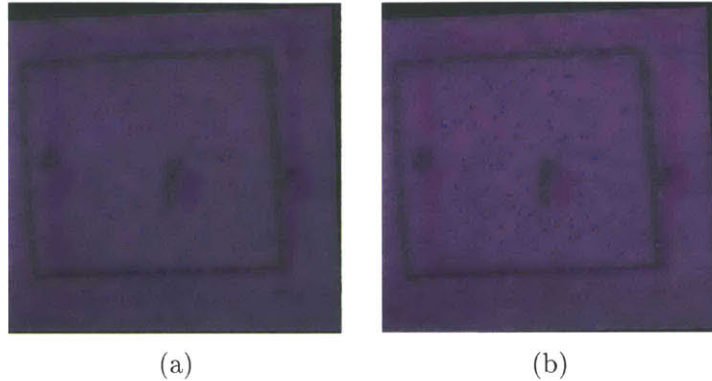
Figure 4-4: Matched keypoints from keypoint matching for the 2 rounds of keypoint matching. Red keypoints represent keypoints from the target image T (Figure 4-3a) and blue keypoints represent keypoints from the source image S (Figure 4-3b). The keypoints are superimposed on the images to denote their location. The green image is T and the red image is the rigidly registered source image S_{H_2}. (a) Matched keypoints (1st round). (b) Matched keypoints (2nd round).

However, such a favorable situation does not always occur because RANSAC does not guarantee that a ground truth model is found within the selected number of trials and the probability of finding inlier keypoints is unknown for different pairs of skin images. Failure cases occur when the 1st round of keypoint matching fails to find a ground truth homography model. Also, the confidence that a homography model is the ground truth increases with the number of matched keypoints found. Thus, identifying factors that may cause failure of the 1st round of keypoint matching is beneficial.

A dataset consisting of a few categories of skin image pairs was made. Images of resolution 1340×1340 from three subjects were used: Subjects R and A are of South Asian descent and subject M is of Caucasian descent. Image pairs were constructed based on different orientations of the skin scene, artificial stretching of skin, taking images over a period of time, comparing skin patches among different persons and different parts of a body part from the same person. The number of correspondences found after the 1st and 2nd rounds of keypoint matching and registration are shown

<p align="center">101</p>

in Table 4.1. Due to a reasonable chance of registration failure in the 1st round, we ran the 1st round on each image pair 5 times and took the highest number of correspondences found. Thus, we define a successful rigidly registered result as one in which an image pair is registered after 5 times, or equivalently after 5000 total extracted minimal sample sets (MSS). The results show that our method is able to register images from different orientations, with many matched keypoints found in the 1st round. However, for a case of upwards translation, much less keypoints were found because of the large curvature of the ventral forearm moving upwards. The homography model also good for handling images with artificial skin stretching in some direction. Furthermore, for time-lapsed images, we found that for most cases, the correct homography model was found. Despite changes in camera settings and illumination settings between image pairs, few but sufficient numbers of matched keypoints were found for registration. This shows that the SIFT descriptor is fairly invariant to illumination differences. These image pairs with few matched keypoints are more likely to require more RANSAC trials due to lower probabilities of finding inlier MSS. In the failure cases, the combination of illumination differences and large translations made those cases challenging even for human observers to estimate the transformation. Finally, when comparing images between different persons or different parts of a body part, we obtained failure, just as expected.

From the above analysis, the factors that cause failure or very few correspondences found in the 1st round are summarized as follows: First, large translations; this causes local illumination changes that affects the appearance of the microrelief. Also, if the curvature of the skin surface is large, rigid registration cannot account for most of the transformation. Second, illumination differences due to different imaging system settings; this reduces the number of matches based on the SIFT descriptors. Third, the density and thickness of hairs; high densities and hair thickness occlude the skin surface and reduces the number of correspondences found.

Another general observation is that for a typical skin image, up to 13000 keypoints are first identified by the SIFT detector. Of these keypoints, the 1st round identifies anywhere from 20 to 2000 inlier matched keypoints. The 2nd round drastically im-

proves the number of matched keypoints identified; the improvement factor is higher if few matched keypoints were identified in the 1st round.

Table 4.1: Number of correspondences found for different image pairs from the 1st and 2nd rounds of keypoint matching and global registration.

Subject	Image Pair Type	Success?	1st Round Correspondences	2nd Round Correspondences
Different Orientations				
R	Rotation of 45 Degrees Anti-Clockwise	Yes	721	921
R	Right Translation	Yes	791	1045
R	Left Translation	Yes	1071	1500
R	Upwards Translation 1	Yes	356	550
R	Upwards Translation 2	Yes	44	192
R	Downwards Translation 1	Yes	251	547
R	Downwards Translation 2	Yes	1010	1500
Artificial Stretching of Skin				
R	Compression (x-axis)	Yes	38	67
R	Outward Stretch	Yes	2058	2509
Different Days with Different Imaging System Settings				
A	Day 1 to Day 49	Yes	31	127
A	Day 1 to Day 147	Yes	23	79
A	Day 49 to Day 147	No	-	-
R	Day 1 to Day 98	No	-	-
R	Day 1 to Day 103	No	-	-
R	Day 1 to Day 107	Yes	21	63
R	Day 98 to Day 103	Yes	67	103
R	Day 98 to Day 107	Yes	35	115
R	Day 103 to Day 107	Yes	52	148
Different Persons				
A/B	A and B	No	-	-
A/R	A and R	No	-	-
A/M	A and M	No	-	-
B/M	B and M	No	-	-
Different Areas of Body Part				
R	Different Areas of Ventral Forearm	No	-	-

Evaluation with Simulated Images

Accurate rigid skin registration within the ROI is essential for accurate microrelief matching. As such, we use ground truth SIFT keypoints from simulated images to verify accuracy. Distorted simulated images I_{sim} are created from a real base image I_{base} using the process detailed in Chapter 3.2.6. Next, the matched keypoints are found between the image pair are globally registered, thus obtaining Image S_{H_2}. As keypoints from I_{sim} are not ground truth, they are discarded. Instead, keypoints from I_{base} are transformed into the domain of I_{sim} and then back-transformed into the domain of S_{H_2} using the homography found by global registration. Denoting the simulated transformation as $\mathbf{T}_{sim}$, the root-mean-square error (RMSE) of the final

set of keypoints K_f in the domain of I_{base} is given by:

$$\text{RMSE} = \sqrt{\frac{1}{|K_f|} \sum_{\mathbf{x} \in K_f} \|\hat{\mathbf{H}} \mathbf{T}_{sim}(\mathbf{x}) - \mathbf{x}\|^2} \qquad (4.6)$$

The base image I_{base} is displayed in Figure 4-5a (an extended 2040×2040 version of Figure 4-3a). For the PPM algorithm to work as desired, it is important to identify a distortion function that yields a single cluster of uniformly distributed points after keypoint matching; this reflects the curvature of the skin surface in 3D space. We showcase three plausible functions for distorting real image I_{base} in Table 4.2. The stretching parameters, camera motion parameters and warp factor are also displayed in the same table. The RMSE and number of matched keypoints for simulated images S1, S2 and S3 can be found in Table 4.3. We can also find the dense error vector field after rigid registration using a dense grid instead of keypoints as in Equation 4.6. The length of a yellow arrow indicates the magnitude of the error and direction of the arrow indicates that a pixel in real image I_{base} needs to move in that direction to reach its true location in distorted image I_{sim}. The simulated images and their corresponding error vector fields are shown in Figure 4-5.

Table 4.2: Parameters for simulated images for testing rigid registration.

Simulated Image Name	$f_x(x)$	$f_y(y)$	θ_s	s_x	s_y	θ_r	θ_p	θ_y	w
S1	$\cos(0.002x - 3\pi/4) + \cos(0.003x)$	$\sin(0.002y + 3\pi/4) + \cos(0.003y)$	10	1.05	1	-10	10	10	5
S2	$\cos(0.005x) + \cos(0.006x)$	$\cos(0.004y) + \sin(0.006y)$	-45	1	1.1	5	5	-10	5
S3	$\sin(0.002x) + \cos(0.003x)$	$-\sin(0.003y) - \cos(0.004y)$	40	0.95	1.1	10	10	0	5

Table 4.3: RMSE and number of correspondences of simulated images.

Simulated Image Name	RMSE	Number of Correspondences
S1	0.3263	1907
S2	0.2701	514
S3	0.3222	578

Therefore, the results from the error vector fields in Figure 4-5 show that local planar region in an image is a region where the mean displacement errors are sub-pixel. At the matched keypoints, the RMSE is small (Table 4.3). Assuming that a
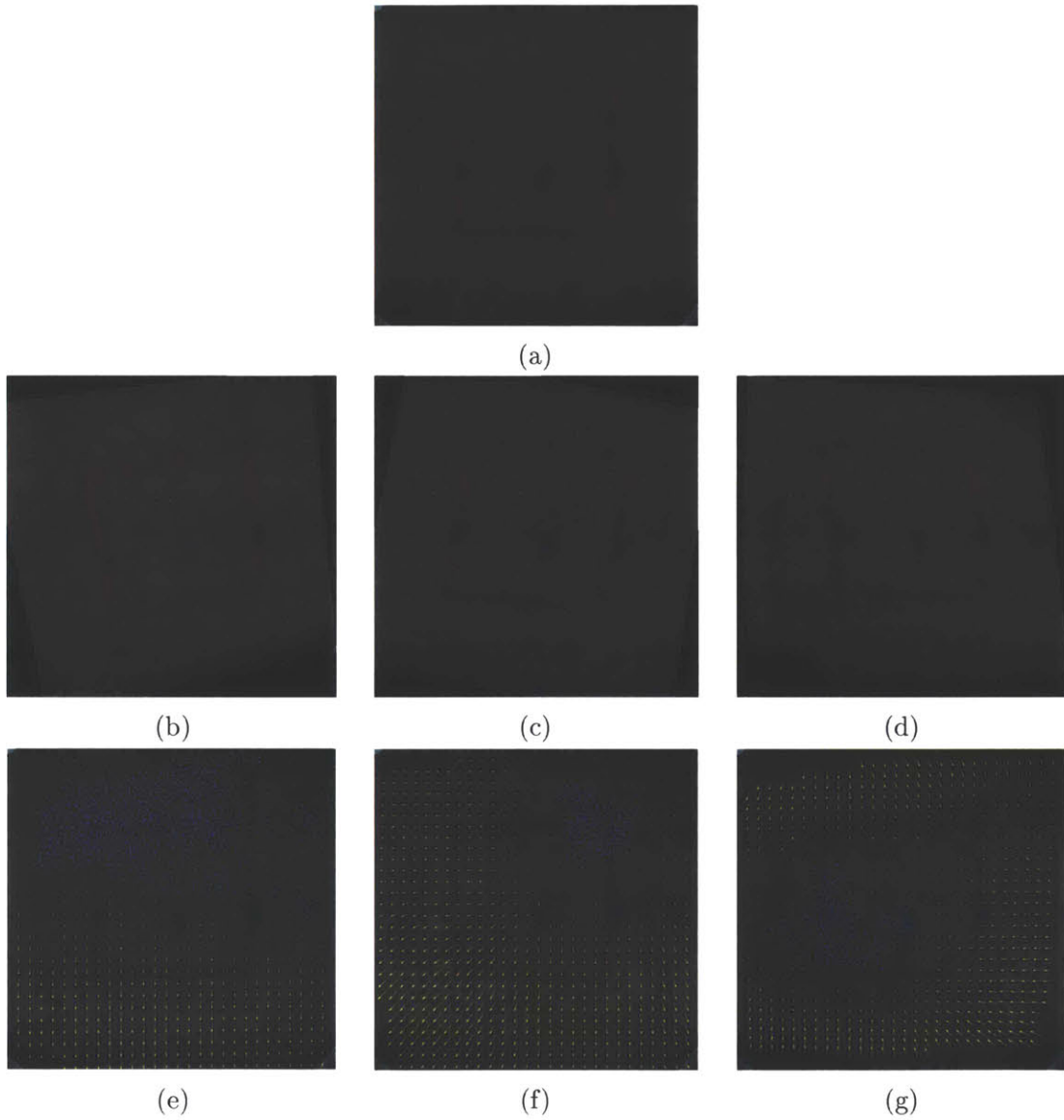
Figure 4-5: Simulated images for global registration and global registration results. The length and direction of the yellow arrows in (e-g) indicate the error vector field and the blue keypoints indicate the true keypoints used in real base image I_{base}. (a) I_{base}. (b) S1. (c) S2. (d) S3. (e) Error deformation field and keypoints for S1. (f) Error deformation field and keypoints for S2. (g) Error deformation field and keypoints for S3.

true matched keypoint is identified, there is some neighborhood around that keypoint where rigid registration is sub-pixel accurate.

The next course of action is to evaluate rigid registration for robustness to various distortion scales or warp factors w and yaw or in-plane camera rotation angles θ_y. We generated different versions of simulated image S3 by varying $w = \{2, 2.5, 3, 3.5, 4, 4.5, 5\}$ and $\theta_y = \{-10°, -5°, 0°, 5°, 10°\}$. The RMSE (Equation 4.6) and number of correspondences found are shown in Figures 4-6a and 4-6b respectively. The RMSE plot indicates that accuracy of feature-based homography registration at the matched keypoints is invariant in-plane rotation angle and is robust to realistic amounts of distortion. Also, RMSE of true keypoints is sub-pixel accurate. Furthermore, the number of correspondances found does not depend on rotation angle, but it decreases as warp factor increases. This observation suggests that smaller local planar regions are found when the surface curvature of the skin is made higher by nonlinear distortions.
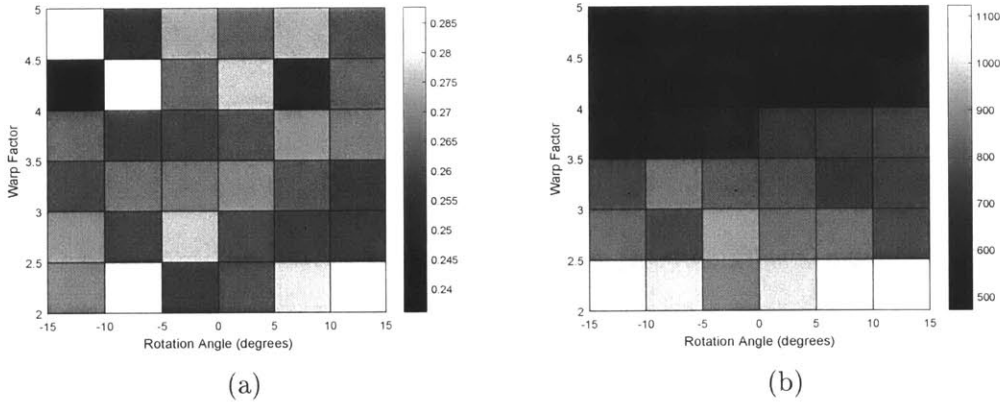
Figure 4-6: RMSE and number of correspondences as a function of warp factor w and yaw θ_y. (a) RMSE. (b) Number of correspondences.

In conclusion, we have sufficient evidence to use matched keypoints from rigid registration as a method for identifying a region of registered ROIs in a skin patch for accurate microrelief matching. The homography transformation is able to account for varying camera angles and in-plane rotations, some amount of skin stretching, a fair amount of illumination differences and nonlinear distortion within the identified ROI. Also, matched keypoints reside in a region of sub-pixel registration RMSE. Hence,

for almost-flat skin image surface with some sparse amounts of hair, there is a good chance that we can apply rigid registration and find a cluster of evenly distributed keypoints for convex hull registration.

4.2.3 Microrelief Matching

Evaluation with Real Images

For demonstration of MJPM, we use the image pair T and S given in Figure 4-3. S is globally registered so that we obtain image S_{H_2}. In Figure 4-7a, the two convex hulls from S and S_{H_2} were extracted from the matched keypoints of a homography transformation between the two images. The convex hulls were determined from the matched keypoints shown in Figure 4-4a. In each skin patch (skin region within the convex hull), the microrelief structure and junction points were extracted (Figure 4-7b, c). The ROI of each skin patch was divided into concentric rings as shown in Figure 4-7d. Junction points in each indexed ring were matched between paired skin patches (Figure 4-7e, f). Figure 4-7g, i illustrates that junction points in the skin patch centers matched well.

Evaluation with Simulated Images

In this section, we evaluate the accuracy of the PPM algorithm in terms of robustness to yaw and local and global distortion. The image pair used is that of the real base image I_{base} and variants of S3 (Figure 4-5d). The microrelief structure of I_{base} is prominent and has relatively high contrast; this makes it a good candidate image for preliminary microrelief matching investigation.

To quantify the accuracy of our PPM algorithm, we used the matching error, defined as the average Euclidean distance in pixels between the ground truth junction points in real image I_{base} and junction points from the matching result.

Different I_{sim} were generated using various in-plane rotation angles $\theta_y = \{-10°,$ $-5°, 0°, 5°, 10°\}$ (keeping warp factor w constant at 5 pixels) and warp factor of $w = \{2, 2.5, 3, 3.5, 4, 4.5, 5\}$ (keeping yaw θ_y constant at 0°). The average and
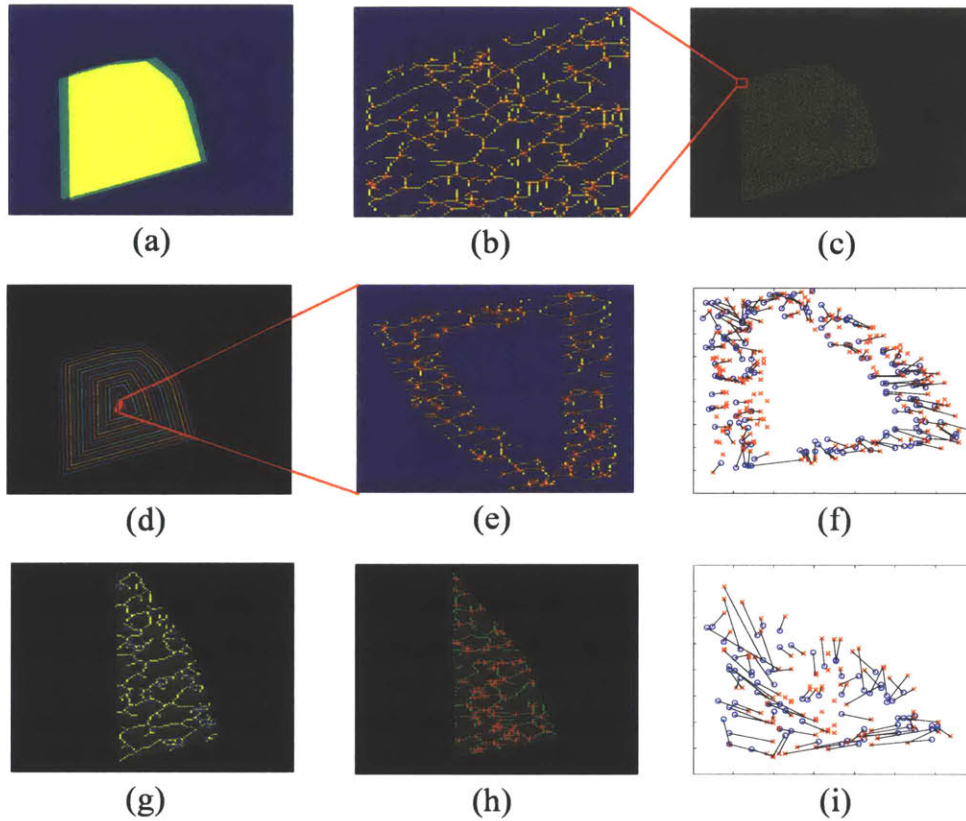
Figure 4-7: Microrelief matching process of skin patches on the forearm of the same subject showing skin patches from images T and S_{H_2} being matched. Target image T was taken on 09/29/2015 and source image S was taken on 02/23/2016, as shown in Figures 4-3a and 4-3b respectively. Then, image S_{H_2} is obtained by homography warping of S. (a) Convex hull registration of paired skin images T and S_{H_2} to obtain paired skin patches. (b) (c) Extracted microrelief structure and junction points of image T. (d) Concentric rings of skin patch from image T (boundaries of rings are plotted) conform to the shape of the convex hull (e) Microrelief structure and junction points of the smallest, innermost ring in image T. (f) Matching result of the pair of smallest rings in S_{H_2} and T (g) Central microrelief structure map in skin patch from image T. (h) Central microrelief structure map from image S_{H_2}. (i) Matching result of the skin patch centers of T and S_{H_2}.

median matching errors are shown in Figure 4-8. For reference, the third index points of the two curves in Figure 4-8a and the seventh index points of the two curves in Figure 4-8b refer the average matching errors corresponding to image pair I_{base} and S3 (Figure 4-5d). All of the average matching errors are within two pixels and all median matching errors are sub-pixel. Thus, the MJPM algorithm is invariant to in-plane rotation angles and is robust to various distortion scales. The median and mean matching error values illustrate that most of the matching errors were below 1 pixel while missing points and extra points contributed in large matching errors, this causing the mean matching error to be inflated.
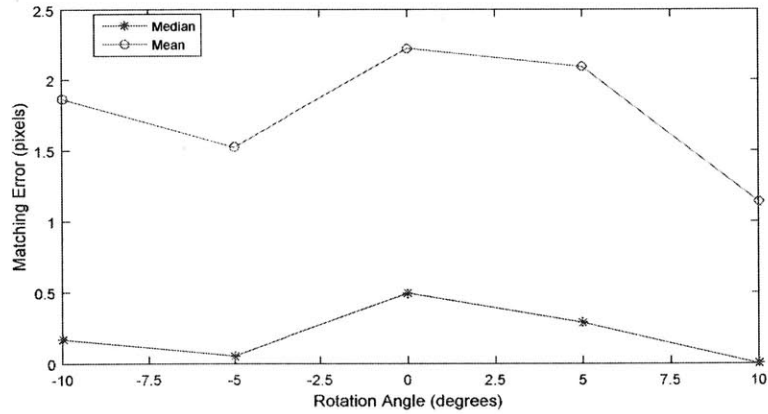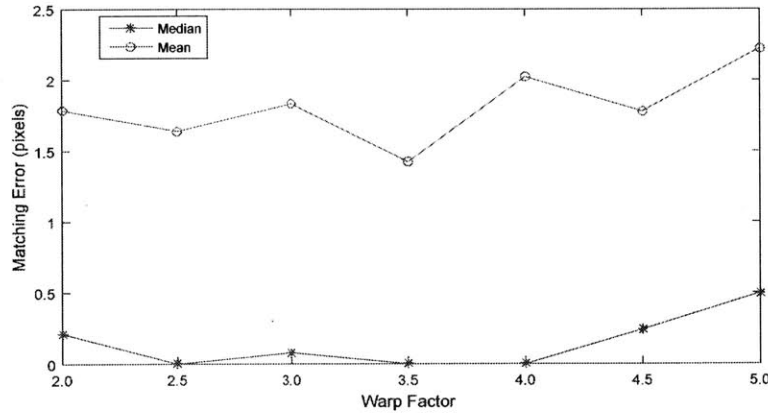
Figure 4-8: Matching Error for skin microrelief matching. (a) Rotation angle and matching error. (b) Warp factor and matching error.

4.3 Discussion

In this chapter, we presented MJPM: a rigid skin registration and a nonrigid PPM framework to monitor the evolution of the skin microrelief structure over time by precisely tracking the movement of microrelief junction points. Our experiments showed that rigid skin registration achieves sub-pixel RMSE in a locally planar ROI on the skin. Then, the registered ROI is further refined by nonrigid PPM, which matches two sets of microrelief junction points. In a proof-of-concept evaluation, the PPM algorithm was shown to exhibit sub-pixel median matching error values and mean matching error values of less than two pixels. It is invariant to in-plane rotation angles and robust to different distortion scales. Thus, we have verified that our algorithm remains accurate and robust in matching pairs of distorted and transformed skin patches. As matching errors are mostly caused by both missing and extra junction points, MJPM would certainly benefit from improved microrelief structure detection and prediction of missing junction points on skin. Also, an evaluation of MJPM on larger datasets with an various skin types and pathological conditions will lead to more conclusive results on the efficacy of this new method for skin image analysis and tracking.

Chapter 5

Hair Segmentation

Hair segmentation is optimal partitioning of pixels in a skin patch image into hair and skin pixel classes. In Chapter 2.4, we discussed that hair detection algorithms in the literature follow the general process of finding potential hairs with curvilinear structure detectors and then filtering out unwanted structures by mathematical morphology. They may also fill in holes among detected hair pixels and gaps between elongated hair using various techniques. In this chapter, we take a step-by-step approach to designing a hair segmentation algorithm by investigating various curvilinear structure detectors and designing an adaptive threshold based on observations of the hair segmentation mask. At each step, we conducted experiments and used our observations to determine the final design.

5.1   Overall Algorithm Design and Implementation

5.1.1   Overall Design

In Figure 5-1, the flowchart of hair segmentation shows four main steps in our algorithm. First, the likelihood of hair-like structures in the input hair-occluded skin image is encoded in a response image. Second, an adaptive threshold gives an initial guess of the hair pixels in the response image to output a hair segmentation mask. Third, a k-nearest neighbor (k-NN) classifier refines the mask by taking hair and skin

colors into account. Fourth, a series of morphological operations filter out non-hair objects from the mask.

Hair-Occluded RGB Image

⬇

| Curvilinear Structure Detection |

⬇

| Adaptive Thresholding |

⬇

| *k*-NN Classification |

⬇

| Morphological Processing |

⬇

Binary Hair Mask
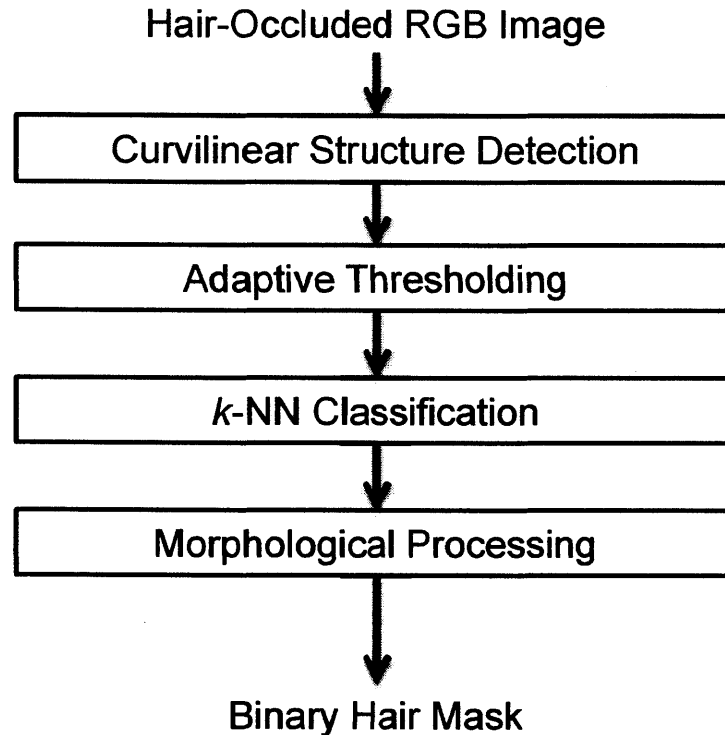
Figure 5-1: Algorithm flowchart for hair segmentation.

5.1.2 Implementation

We implemented our hair detection algorithm on MATLAB Release 2015b on a system with an Intel Xeon E3-1225 v3 3.2 GHz processor. As there are no public implementations available for many published digital hair removal (DHR) methods, we implemented our own versions of some these algorithms on MATLAB based on written description and default paramaters in the literature. The algorithms we implemented were devised by Abbas et al. [3], Fiorese et al. [21] and Xie et al. [104]. The only publicly-released DHR method tested is DullRazor [53]. The implementation can be found at [2].

5.1.3 Datasets

Using our imaging system, a set of 15 hair-occluded skin images of resolution 400×400 were acquired. The hairs were hand-annotated to serve as ground truth. The images cover various skin (from dark to light skin tones), hair types (different lengths, thicknesses and colors) and illumination settings. In most images, microrelief structure is visible beneath the hairs. The subjects were of Caucasian, South Asian and East Asian descent. Most existing algorithms were not developed for detecting light hairs; we did not include such images in this dataset. The images shown in Figure 5-2 were used for algorithm comparison.
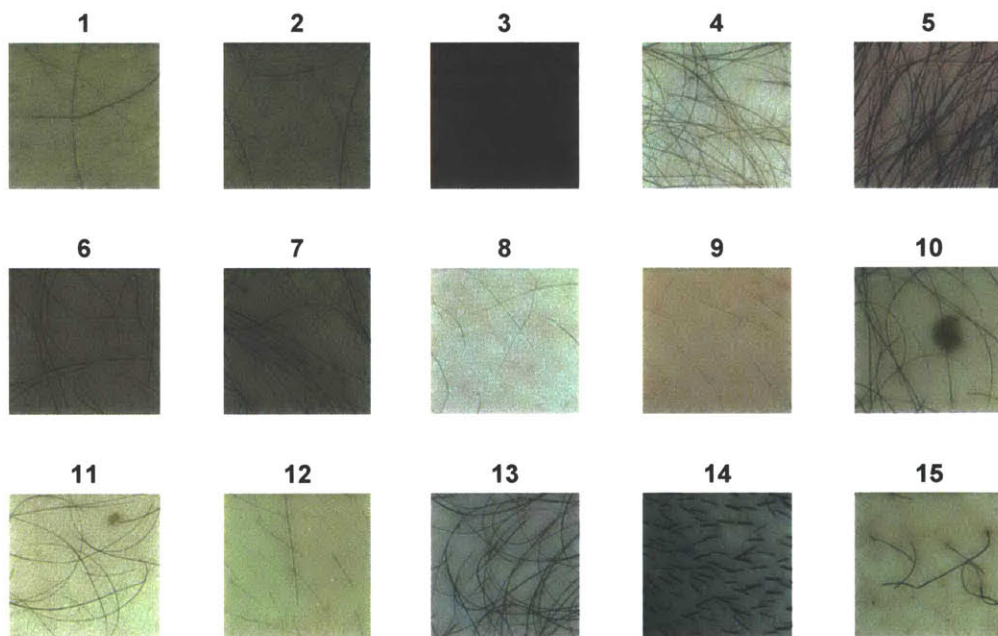
Figure 5-2: Hair-occluded skin image dataset used for algorithm comparison. The images are numbered for referencing. The hairs were hand-annotated for ground truth (not shown).

It is also useful to test our method on skin images without hair. In these images, the microrelief structure in these images are clearly visible and the skin tone varies from dark to light skin. This helps verify that our hair segmentation method is able to identify that a skin image is without hair and outputs the correct all-false hair segmentation mask. The images in this dataset (resolution 400×400) are shown in Figure 5-3.
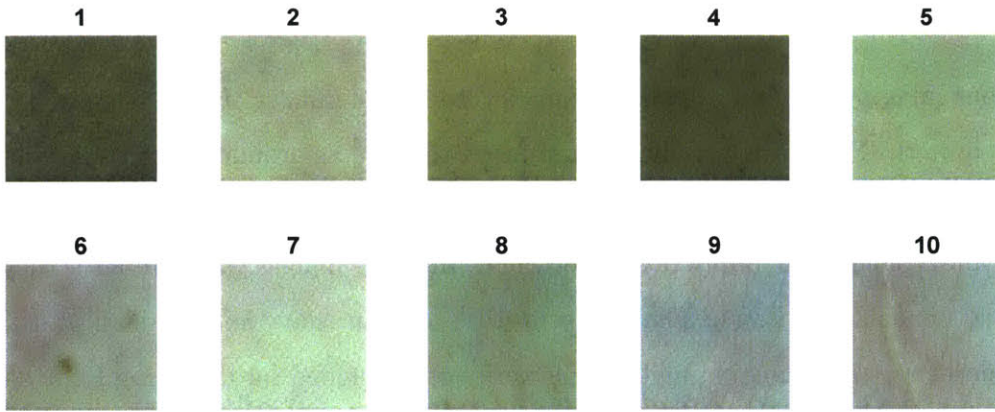
Figure 5-3: Dataset for skin-only images.

In addition, to test whether our method works on the dorminant interest in the literature, a dataset was created out of 10 dermoscope images with pigmented skin lesions (PSL) and dark hairs used by Koehoorn et al. [46] in their hair segmentation evaluation. The images in this data (resolution 1024×768) are shown in Figure 5-4.
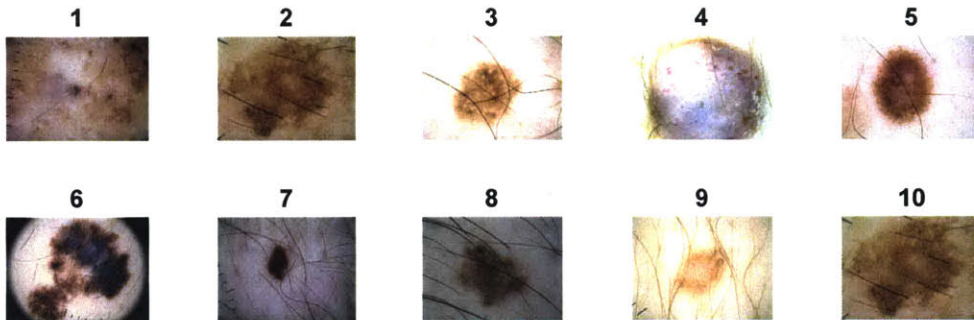
Figure 5-4: Dermoscope image dataset showing images with PSLs and dark hairs.

5.2 Curvilinear Structure Detection

5.2.1 Exploring Color Spaces

Although hair-occluded skin images are acquired in RGB color space, it is much easier to design curvilinear structure detection for single-channel images. It is desirable to preserve color information from RGB images so that there is sufficient contrast in image intensities between true hair and skin pixels. It is also advantageous to

114

de-emphasize pixels belonging to the microrelief structure. We conducted a visual comparison of various color spaces on two potentially challenging hair segmentation tasks: (1) dark hairs with high contrast skin microrelief (Figure 5-5a); (2) mixture of dark and light hairs (Figure 5-5b). We display the red, blue and green channels, L component of Lab, Y component of YUV (BT.601), V component of HSV, 1st component of principal component analysis (PCA) [102], 1st component of JADE independent component analysis (ICA) [13] and 1st component of FastICA [34] for both images in Figures 5-6 and 5-7 respectively. The rationale for using PCA and ICA as color spaces is that the new color space would be adapted to colors in the image according independent or uncorrelated color vectors respectively [56], thus potentially making skin and hair colors more distinct. To use the PCA and ICA algorithms, the intensity values from the 3-channel RGB image are rearranged into a $3 \times N_{pixels}$ data matrix, where N_{pixels} is the number of pixels in the RGB image. The output $3 \times N_{pixels}$ matrix is then rearranged back into a 3-channel image.

For dark hairs with microrelief in Figure 5-6, the red channel and 1st component of JADE color space could been seen to de-emphasize the microrelief on the skin, while enhancing the dark hairs. In the remainder of the images, the microrelief were preserved to various extents. Comparing the red channel and 1st component of JADE in terms of computations, the red channel requires no computations whereas JADE requires expensive computation of independent components, which does not scale well with higher number of pixels. Therefore, the red channel is most suitable for increasing sensitivity of curvilinear structure detectors to hair colors, while decreasing their sensitivity to skin and microrelief colors. However, the red channel is prone to demosaicing artifacts of the camera sensor in images in which the dark hairs exhibit high contrast with the skin. It would be advantageous to reduce these artifacts while preserving the benefits of using the red channel.

In Figure 5-7, we observe that the green channel and 1st components from Lab, YUV and PCA color spaces were effective in preserving the contrast of light hairs with respect to the skin background. However, these color channels also preserve the microrelief structure, which is darker than the background skin. As a curvilinear

structure detector can be set to be sensitive to light objects on dark background, the dark microrelief structure will have a low response while light hairs will have a high response. Thus, the microrelief structures can be assumed not to interfere with light hair detection. Among the receptive color channels, we selected the single-channel images with the least demosaicing artifacts: an artifact that is apparent for thin hairs. Although there is no perceptible difference between the image components in YUV and Lab, the Y component in YUV is a simple linear transformation of the RGB image as opposed to the nonlinear Lab transformation. Therefore, the Y component of YUV is the best color channel for light hair detection.

(a)   (b)
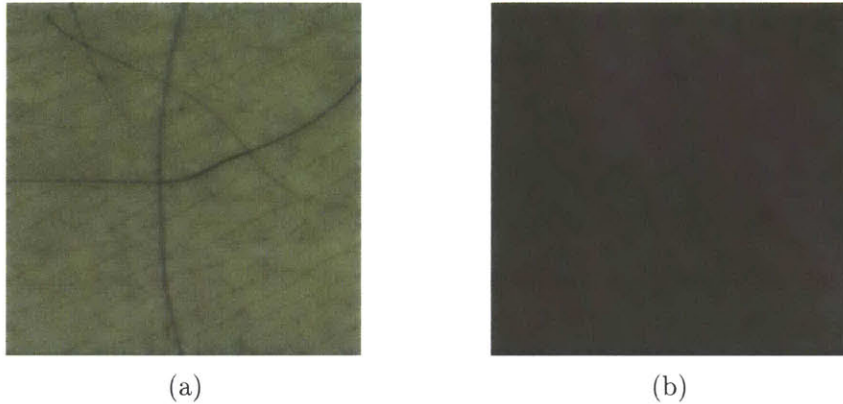
Figure 5-5: Test images for hair segmentation demonstration. (a) Dark hairs with high contrast skin microrelief. (b) Mixture of dark and light hairs.

5.2.2   Top-Hat Transform

Let I_{bw} be the single-channel image of skin with hairs. In order to emphasize dark objects on a light background and correct for uneven background illumination, the black top-hat transform of I_{bw} subtracts I_{bw} from its closing:

$$I_{th} = |\max_i(I_{bw} \bullet b_i) - I_{bw}| \tag{5.1}$$

where $\bullet$ is the grayscale morphological closing operator [91] and structuring elements (SE) b_i are one-pixel width lines of length 13 at the following angles $\{0°, 45°, 90°, 135°\}$. The SEs are flat with unit height, symmetrical and have their origin at the center.

Red Channel **Green Channel** **Blue Channel**

Lab (L Component) **YUV (Y Component)** **HSV (V Component)**

PCA (1st Component) **JADE (1st Component)** **FastICA (1st Component)**
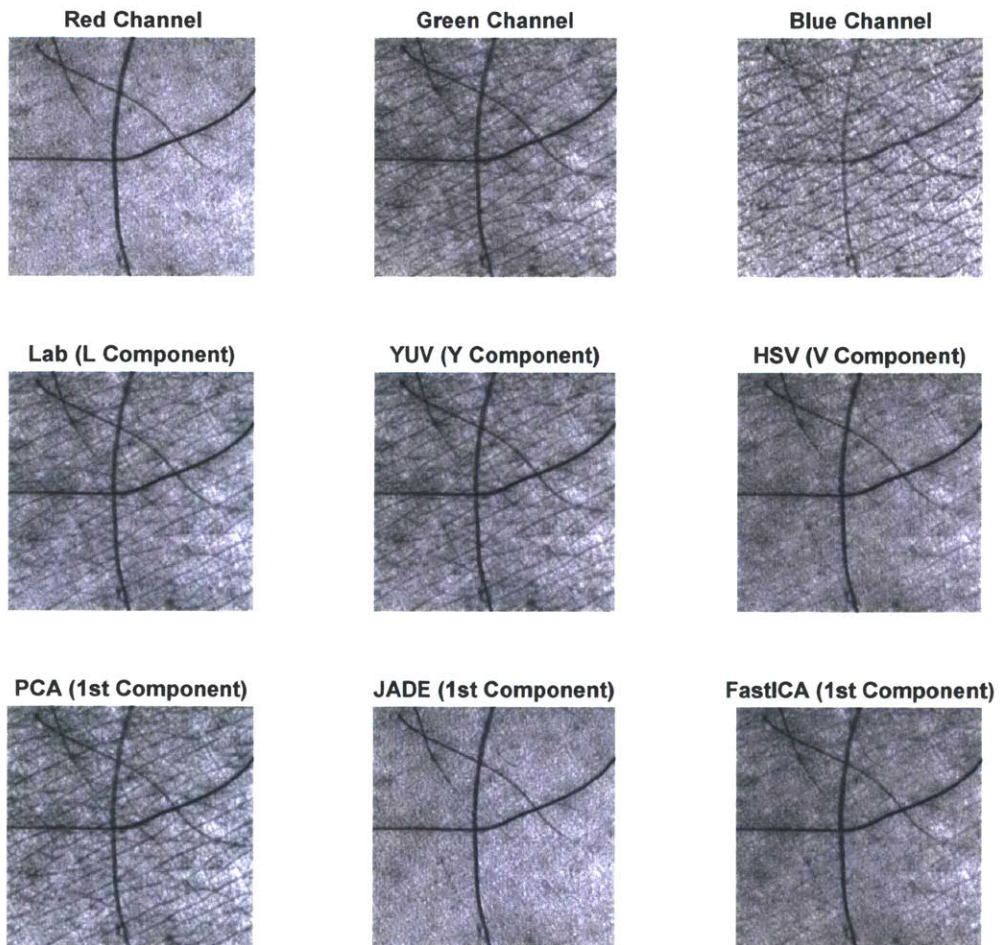
Figure 5-6: Single-channel images for dark hair with microrelief.

Red Channel   **Green Channel**   **Blue Channel**

Lab (L Component)   **YUV (Y Component)**   **HSV (V Component)**

PCA (1st Component)   **JADE (1st Component)**   **FastICA (1st Component)**
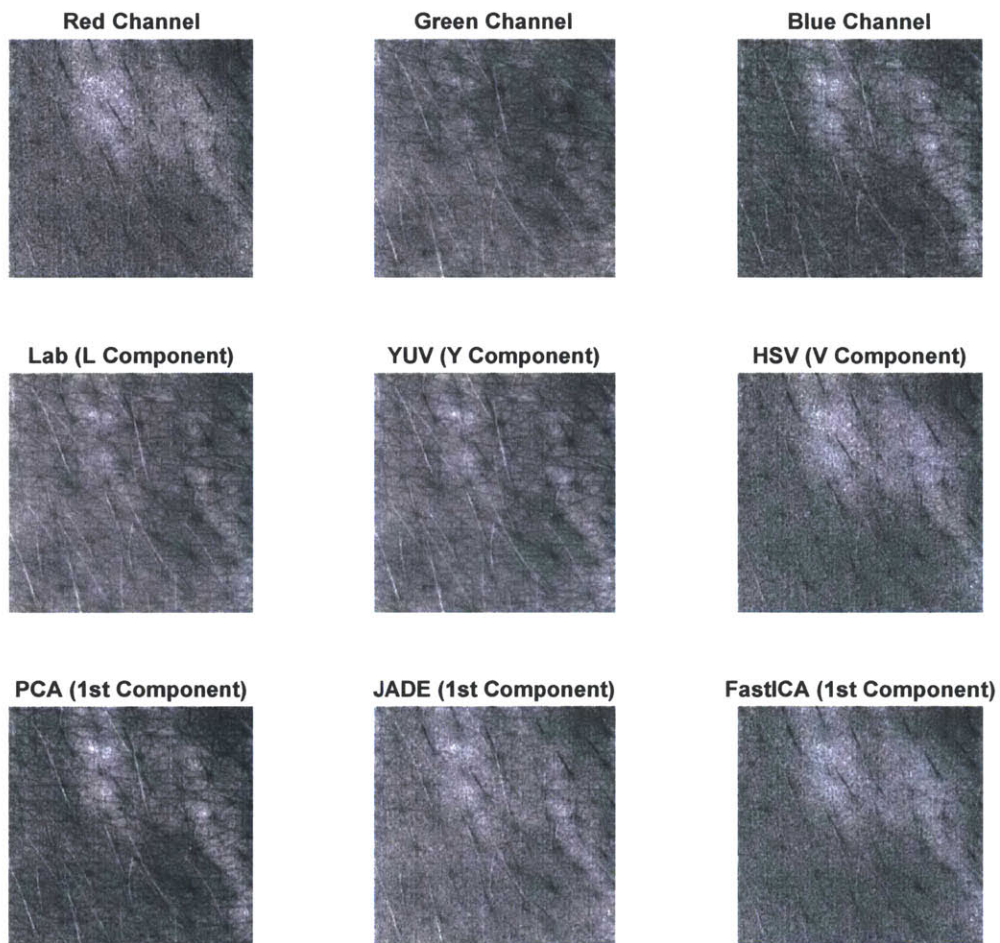
Figure 5-7: Single-channel images for mixture of dark and light hair.

They are chosen in order to preserve dark line-like structures in I_{bw}.

Similarly, in order to emphasize light objects on a dark background (i.e. light hairs), the white top-hat transform of I_{bw} subtracts the opening of I_{bw} from I_{bw}:

$$I_{th} = |I_{bw} - \max_i(I_{bw} \circ b_i)| \qquad (5.2)$$

where $\circ$ is the grayscale morphological opening operator.

Figure 5-8 shows an example of the top-hat transform applied on the mixture of dark and light hairs image shown in Figure 5-5b. The red channel is used for the black top-hat transform and reveals dark hairs in the image, while the white top-hat transform, using the Y component of YUV, reveals light hairs in the same image.

(a)   (b)
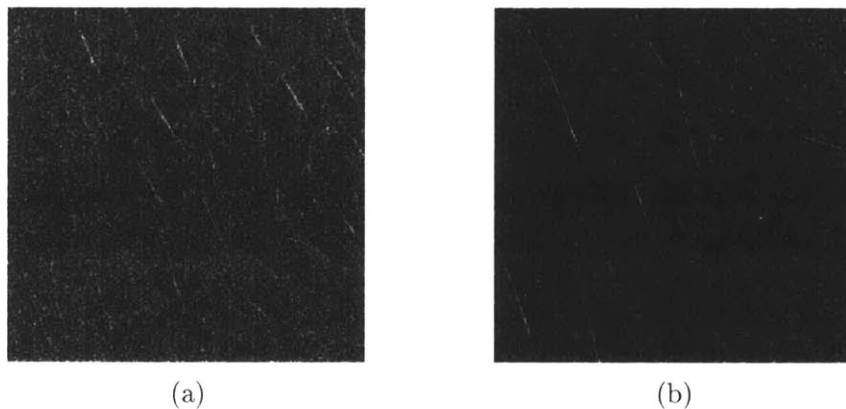
Figure 5-8: Top-hat transform on a test image of a mixture of dark and light hairs shown in Figure 5-5b. (a) Black top-hat transform reveals dark hairs. (b) White top-hat transform reveals light hairs.

5.2.3 Multiscale Matched Filter

Next, a multiscale matched filter for curvilinear structure detection is applied on I_{th}. As explained in Chapter 2.4, several assumptions underpin the rationale for matched filtering. Matched filtering assumes that hair has a small curvature and can be approximated by piecewise linear segments. It also assumes that hair has a cross-section intensity of a Gaussian added to a base background intensity (see Figure 2-8). For our case, we add the assumption that the properties from the above assumptions

are preserved after the top-hat transform. Following work in [26] and [88] for retinal vessel detection, we employ an amplitude-modified second-order Gaussian filter for multiscale hair detection.

The second-derivative Gaussian filter is given by:

$$G''(x) = \frac{x^2 - \sigma^2}{\sqrt{2\pi}\sigma^t}e^{-x^2/2\sigma^2} \tag{5.3}$$

where the standard deviation σ controls the width or scale of the filter and t is the exponent of the σ term in the denominator of the normalization factor. In the following analysis, by setting an appropriate t, we show that we can detect hair at different widths by normalizing the response of the filter at different filter scales σ [26]. The response of the filter across various σ can then by easily combined in a convenient way such as taking the maximum response at a particular scale. Given a hair Gaussian profile $g(x) = A\exp(-x^2/2s^2)$ with scale s and amplitude $A = -1$ for convenience, we computed the response of the filter in Equation 5.3 to the Gaussian profile across different filter widths σ by convolution. For the standard amplitude second derivative of Gaussian, we set $t = 5$; this is derived by taking the taking the second derivative of the normalized Gaussian:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2} \tag{5.4}$$

The results for the maximum response of the convolution is shown in the dotted line plots in Figure 5-9. The plots show that the amplitude of the response generally decreases as σ increases. This is not desirable because it implies that there is no easy way to compare the response across different σ.

Since the convolution reaches its maximum when the maximum response of the hair Gaussian profile and the Gaussian filter meet, the convolutional peak $c(t)$ is given by:

$$c(t) = -\int_{-\infty}^{+\infty} \frac{x^2 - \sigma^2}{\sqrt{2\pi}\sigma^t}e^{-x^2/2\sigma^2}e^{-x^2/2s^2}dx \tag{5.5}$$

It can be shown that the equation simplifies to:

$$c(t) = \frac{s\sigma^{5-t}}{(s^2 + \sigma^2)^{1.5}} \tag{5.6}$$

An interesting case occurs when setting $t = 3.5$. At this value, for a constant hair Gaussian profile scale $s = s_0$, the convolutional peak reaches a maximum when $\sigma = s_0$. In other words, the maximum response is obtained when the filter scale equals the hair Gaussian profile scale. Thus, from Equation 5.6, we obtain:

$$c(3.5) = \frac{s\sigma^{1.5}}{(s^2 + \sigma^2)^{1.5}} \tag{5.7}$$

The full derivation can be found in [26], but in their analysis, Equation 5.5 is missing the negative sign. We plot the result for the modified amplitude second-derivative Gaussian filter in Figure 5-9. The plots show that for a given hair scale s, a matching filter width σ results in the maximum response. Therefore, normalization allows the per-pixel response to be easily compared at different filter scales by taking the maximum.
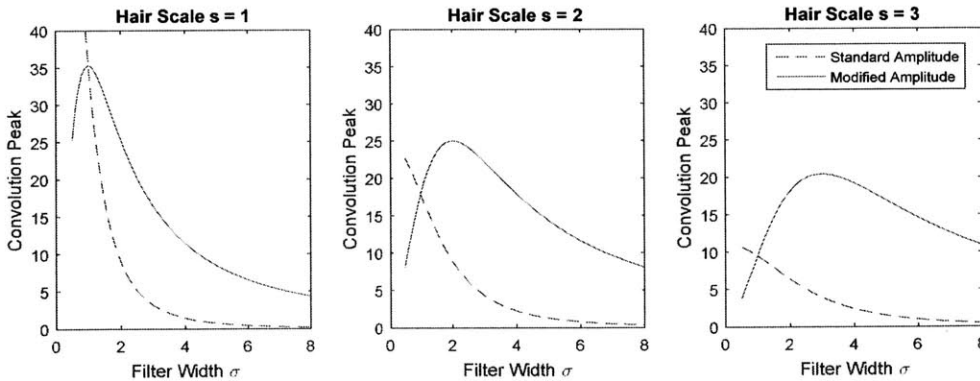
Figure 5-9: Values of convolutional peaks for standard amplitude and modified second derivative of Gaussian while varying the filter scale σ and the hair Gaussian profile scale s, where the profile is given by $g(x) = \exp(-x^2/2s^2)$.

In 2D, the matched filter is defined as:

$$MF(x,y) = -\frac{x^2 - \sigma^2}{\sqrt{2\pi}\sigma^{3.5}}e^{-x^2/2\sigma^2} \qquad \text{for } |x| \leq \nu\sigma, \ |y| \leq L/2 \qquad (5.8)$$

where L is the length of the neighborhood pixels along the y-axis and ν is a constant set at 3 denoting the number of standard deviations from the center pixel. At one filter scale, the matched filter response is the maximum of the responses over a discrete set of filter orientations. Rotations with angle θ are performed with $MF^\theta(x',y') = MF(x,y)$, where $x' = x\cos\theta + y\sin\theta$ and $y' = y\cos\theta - x\sin\theta$. Therefore, the response is simply the maximum response image computed by convolution of I_{th} over all scales and orientations.

We set $L = 13$. A small value of L can detect lines with greater curvature, but the line response is noisy. A large value of L cannot detect lines with large curvature, but the line response is smoother.

For detecting hairs, we choose Gaussian filter scales $\sigma = \{1, 2, 3\}$, which exhibit their maximum response to hairs with Gaussian profile scale $s = \{1, 2, 3\}$, as shown in Figure 5-9. This allows the filter to have a high response to hairs of a range of widths from 1 to 7 pixels, corresponding to widths between 26.6 μm to 186.2 μm, using the assumption that the imaging system resolves skin features that are at least 26.6 μm in length.

12 filter orientations were used, each 15° apart ($\theta = \{0°, 15°, \ldots, 165°\}$. This provides a good balance between having high computational cost from convolving many filters with a single image and having an overall lower filter response due to low resolution of filter orientations.

Therefore, our multiscale matched filter requires convolving 36 filters in total with the top-hat transform image I_{th} and taking the value of the maximum response at each pixel.

5.2.4 Increasing Color Sensitivity

In order to increase sensitivity of our multiscale matched filter to hair colors, we apply the following procedure: First, apply the top-hat transform and multiscale matched filter to both the red channel and Y component of YUV of the input image. Second, the red channel response I_r^R and Y component response I_r^Y are combined by pixel-wise multiplication:

$$I_r(x,y) = I_r^R(x,y) I_r^Y(x,y) \tag{5.9}$$

where x and y are pixel locations. Third, the response I_r is normalized to range $[0,1]$.

Figure 5-8 shows the response of the multiscale matched filter applied on the top-hat transform of the mixture of dark and light hairs image shown in Figure 5-5b. Hair pixels gave a high response, while response of the skin microrelief pixels remained low. The detector is designed so that the user can choose whether to extract dark hairs, light hairs, or both dark and light hairs.
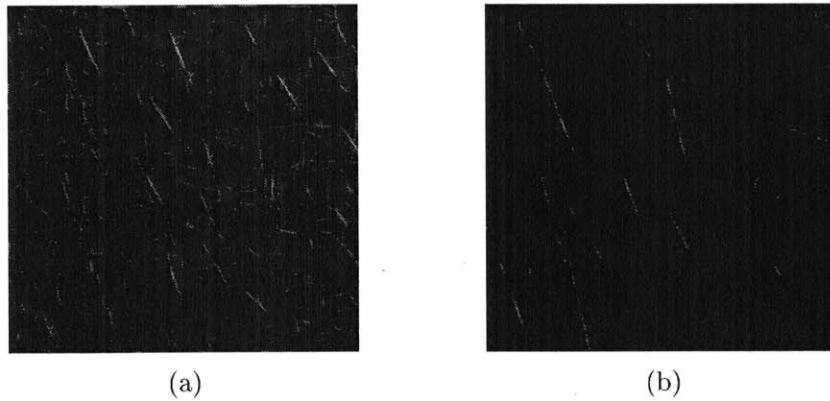

| (a) | (b) |

Figure 5-10: Multiscale matched filter on test image of a mixture of dark and light hairs shown in Figure 5-5b. (a) Multiscale matched filter on black top-hat transform image reveals dark hairs. (b) Multiscale matched filter on white top-hat transform image reveals light hairs.

5.2.5 Comparison with Other Detection Methods

We compared our curvilinear structure detector with three other detectors used in the literature: matched filter (MF) from matched filter-first derivative of Gaussian (MF-FDOG) by Abbas et al. [3], top-hat transform with line structuring elements by Fiorese et al. [21] (similar to the method described in Chapter 5.2.2) and isotropic nonlinear detector (INF) by Xie et al. [104]. For these competing methods, the input grayscale image as is obtained using the Y component of YUV.

The outcome of hair segmentation is a pixel-based classification result. Pixels are labeled either "skin" or "hair". As a result, there are 4 possibilities for a pixel: true positive (TP), false positive (FP), true negative (TN) or false negative (FN). The true positive rate (TPR), or sensitivity, and false positive rate (FPR), given by (1−specificity), are defined as follows:

$$TPR = \frac{\sum TP}{\sum TP + \sum FN} \tag{5.10}$$

$$FPR = 1 - \frac{\sum TN}{\sum TN + \sum FP} \tag{5.11}$$

The receiver operating characteristic (ROC) allows us to evaluate the performance of these detectors as the discrimination threshold of their responses are varied. For a given response image I_r, we varied the threshold t_r and compared the resulting hair segmentation mask with the ground truth hair mask from the hair-occluded image dataset in Figure 5-2. For each curvilinear structure detector, ROC curves for all 15 skin images were calculated, and then the average ROC curve for each detector was plotted (Figure 5-11). The average ROC curve is computed by taking the mean of the TPR from all data as the FPR is varied. The curves show that our method outperforms top-hat transform, MF-FDOG and INF for a large range of thresholds because its TPR values are higher for every FPR value.

Next, we calculated the area under the curve (AUC) by trapezoidal approximation in order to examine the overall performance. The results, which are shown in Table 5.1, reiterates that our detector performs better than the top-hat transform, MF-
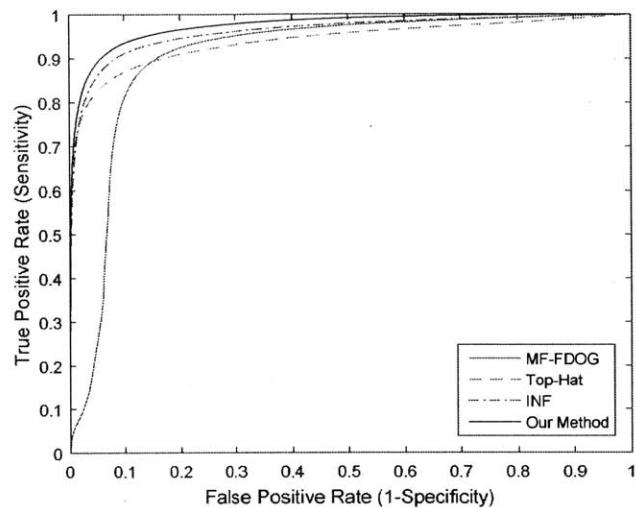
Figure 5-11: Average ROC curves for various curvilinear structure detectors for hair-occluded dataset in Figure 5-2.

Table 5.1: AUC of average ROC for various curvilinear structure detectors on hair-occluded skin image dataset.

Method	AUC
MF-FDOG	0.9054
Top-Hat	0.9391
INF	0.9596
Our Method	0.9735

FDOG and INF. To reinforce our point, we display the response of each detector on the image with prominent microrelief shown in Figure 5-5a. The response image from the top-hat transform (Figure 5-12b) shows that gaps exist in the hair structure. Unlike the top-hat transform, our method is able to automatically fill in gaps in the hair structure. And unlike the MF-FDOG, our method incorporates detectors at multiple hair width scales to better differentiate thin hairs from the background. Moreover, comparing the response of our method in Figure 5-12d and the response of MF-FDOG in Figure 5-12a, our method better emphasizes the hair structures and reflects their true scale with better accuracy. Although the INF performs better than the other 2 competing methods by AUC, the AUC does not reflect whether the structures detected are hair-like. Individual hairs in the hair mask should have smooth outlines and approximately constant width throughout its length. Since the INF uses a circular region of support, it equally enhances small objects and wide lines alike. In Figure 5-12c, tiny dark blobs on the skin induce a high response for the INF. These objects may be adjacent to hair pixels and thus affect the smooth outline of hairs. In contrast, the response for our method (Figure 5-12d) shows that the hair structures are smooth and hair pixels exhibit much higher responses than those of pixels belonging to the skin and microrelief structure.

5.3 Adaptive Thresholding

Finding the appropriate threshold for response image Ir is very important for accurate segmentation. Hair detection methods in the literature use various thresholding methods: the MF-FDOG uses the FDOG as multiplicative pixel-wise weights on a fixed threshold; Fiorese et al. employ Otsu's method [76] to threshold the top-hat transform; the INF is thresholded based on learned parameters of the test dataset from linear regression. These thresholding methods are optimized for detecting high-contrast, dark hairs, and thus do not work well for a handful of our dataset images. In this section, we introduce an adaptive thresholding method for our proposed curvilinear structure detector. If light and dark hairs are chosen to be detected from the same
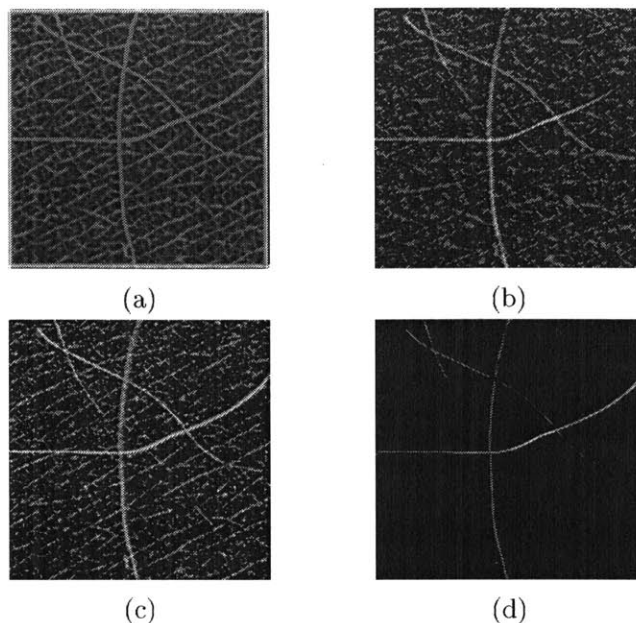
Figure 5-12: Response images from various curvilinear structure detectors. (a) MF-FDOG. (b) Top-hat transform. (c) INF. (d) Our method.

image, then adaptive thresholding is applied separately: one threshold for detecting light hairs and another threshold for detecting dark hairs, and both output masks are combined into a single mask by a pixel-wise OR operation.

5.3.1 Observations

For the 15 images in our hair-occluded skin image dataset, we varied the threshold t_r for 1000 discrete values in range $t_r = [0, 0.999]$ and evaluated the resulting set of segmentation hair masks $\{\mathcal{M}_1, \ldots, \mathcal{M}_{1000}\}$ using functions that reduce each mask into a scalar response value. As such, as t_r is varied, the response value is characterized. And we used observations this relationship to possibly select an optimal threshold.

The first function is the edge density of the mask, denoted by $\mathrm{ED}(t_r)$. First, the gradient magnitude (by forward difference) of mask $\mathcal{M}_i$ is evaluated. Second, the gradient magnitude is normalized by $\sqrt{2}$ division. Third, the mean of all gradient magnitude values is computed. By evaluating all masks $\{\mathcal{M}_1, \ldots, \mathcal{M}_{1000}\}$, values of the function $\mathrm{ED}(t_r)$ are obtained. This function gives a simple measure of the

amount of edge variation in the hair mask with respect to threshold t_r. A high amount indicates that the mask contains a high density of curvilinear structures such as hairs and microrelief, and vice versa.

The second function is the mean branch length of the mask, denoted by $\text{MBL}(t_r)$. First, the skeleton of the mask is obtained using morphological thinning. Second, branch points of the skeleton are removed to reveal unconnected, single pixel width branches. Third, the mean area of the branches is computed. By evaluating all masks $\{\mathcal{M}_1, \ldots, \mathcal{M}_{1000}\}$, values of the function $\text{MBL}(t_r)$ are obtained. This function indicates the average length between branches in the skeleton image with respect to threshold t_r. A low MBL indicates the presence of highly branched microrelief structure while a high MBL indicates that most of the extracted structures are long and hair-like, since it exhibits significantly less branching.

We illustrate both functions in the plot in Figure 5-13; t_r was varied to obtain masks from the image with prominent microreliefs in Figure 5-5a. In this example, at $t_r = 0.008$, ED is at the maximum, where all curvilinear structures in the skin image are extracted. At a particular threshold $t_{r0} = 0.053$, where $\text{MBL} = 10$, a mixture of mostly hair pixels and some microrelief structure were extracted. Misclassified hair pixels did not appear like hair-like structures and could be easily removed by morphological operations. From $t_r = 0.053$ to $t_r = 0.14$, microrelief quickly disappeared but at the same time, hair structures slowly reduced in width, and eventually, gaps appeared in the hair structures. In summary, as t_r increased, misclassified hair pixels increasingly disappeared but misclassified skin pixels increasingly appeared.

In general, for all 15 hair-occluded skin images, the function plots $\text{ED}(t_r)$ and $\text{MDL}(t_r)$ exhibited similar behavior. In ED, the maxima appeared within $t_r = [0.001, 0.01]$. Upon increasing t_r after the maximum point, a characteristic L-shaped appeared in the curve. Values of t_r above the value of t_r at the L-shaped "corner" of the curve indicated a region of stability where many hair pixels remain but most skin pixels are removed. ED then steadily decreased as t_r increased. However, from ED, it is not clear where the optimal threshold lies. For the plot of $\text{MDL}(t_r)$, a peak which may exceed $\text{MBL} = 10$ sometimes appeared at low values of t_r before the function de-

creased to a value below 10 at the value of t_r where ED hit the maximum. Thereafter, $\text{MDL}(t_r)$ was a piecewise monotonically increasing function until $\text{MDL}(t_r) > 10$.

$t_r = 0.008 \qquad t_r = 0.053 \qquad t_r = 0.1 \qquad t_r = 0.14$
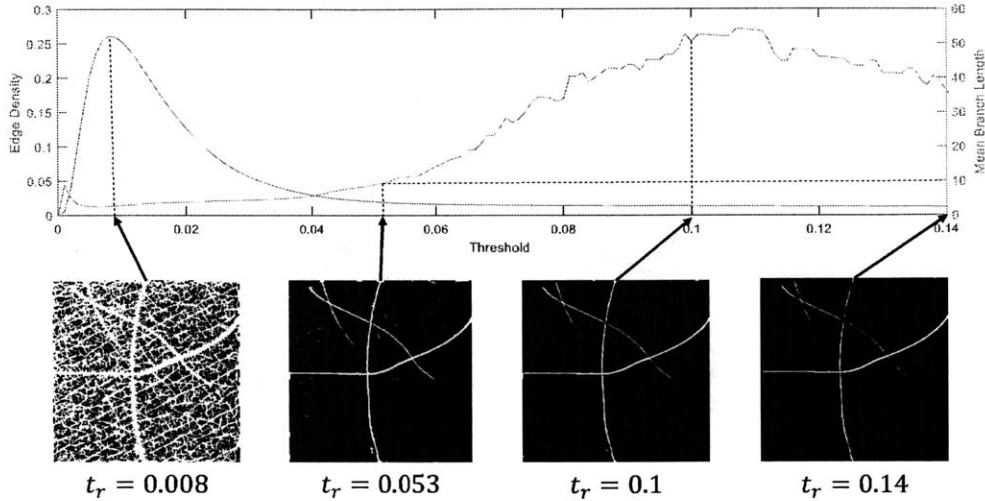
Figure 5-13: Graph of ED and MBL against threshold values and corresponding masks at selected threshold values t_r. Skin image from Figure 5-5a.

Interestingly, for all 15 images in our hair-occluded skin image dataset, when we set MBL = 10, we obtained hair masks at a particular threshold t_{r0} with a mixture of mostly true hair pixels and skin pixels. Even at hair intersections, few gaps appeared in the hair pixels, and the width of the hair was preserved. Most of the hair structures had smooth edges and noisy non-hair structures were mostly non-adjacent to the hair pixels. We display the hair masks of some of these images in Figure 5-14. Overall, the hair masks appeared favorable as initial segmentation masks; simple morphological operations are able to remove non-hair like objects as such the round object appearing in the mask from Image 6 (Figure 5-14a).

To further investigate the above empirical result, we repeated the above procedure on the dataset of skin-only images (Figure 5-3). The function plots $ED(t_r)$ and $\text{MDL}(t_r)$ looked similar to that shown in Figure 5-15, corresponding to image 4 in Figure 5-3. $ED(t_r)$ peaked at values of $t_r = [0.038, 0.079]$. Also, the curve after the maximum was not L-shaped, but instead, $ED(t_r)$ decreased steadily as t_r was increased. The range of the maxima of $ED(t_r)$ for the skin-only dataset did not intersect with the corresponding range in the hair-occluded skin dataset. More notably,
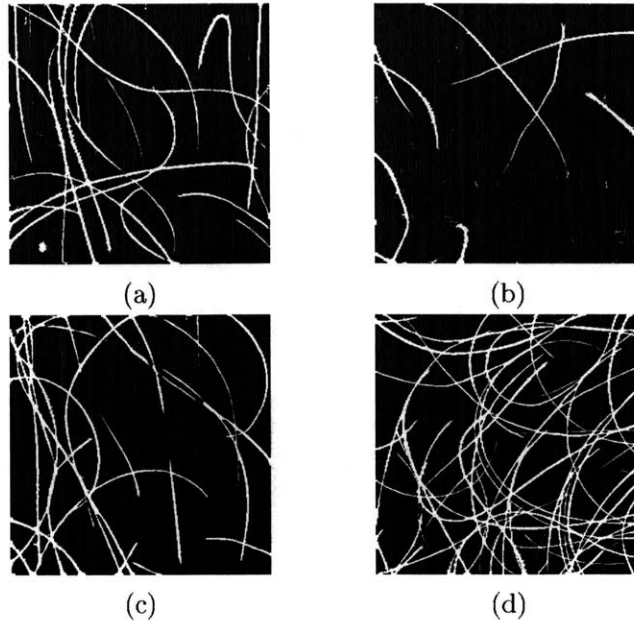
Figure 5-14: Hair masks after thresholding at MBL = 10 for four images taken from dataset in Figure 5-2. (a) Image 6. (b) Image 8. (c) Image 10. (d) Image 13.

after the initial early high peak in $\text{MDL}(t_r)$, the value remained well below 10 as t_r continued rising.
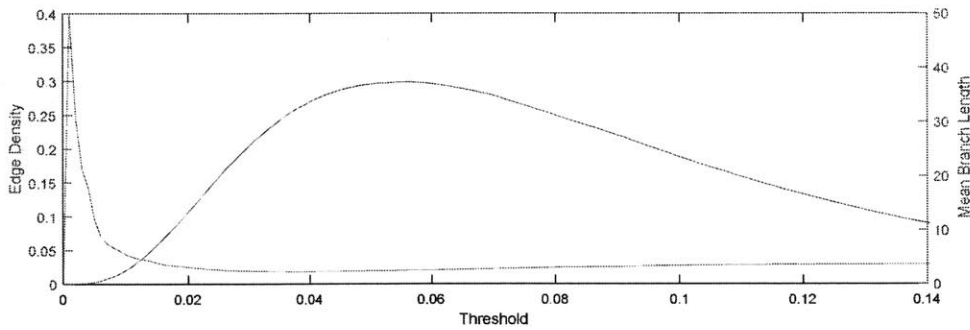
Figure 5-15: Graph of ED and MBL against threshold values for skin-only image. Skin image from image 4 in Figure 5-3.

5.3.2 Threshold Searching

We construct our adaptive threshold based on line search. A single parameter, threshold t_r, is varied to determine an optimal mask image M_o at optimal threshold $t_r 0$ from response image I_r.

To find the optimal threshold, first a threshold t_{max} is found by solving a problem specified by:

$$\underset{0<t_r<0.14}{\arg\max}\ \mathrm{ED}(t_r) \qquad\qquad (5.12)$$

The problem is solved by golden section and parabolic interpolation line search [23].

Second, find the optimal threshold t_{r0} at $\mathrm{MBL}(t_r) = 10$ in range $t_{max} < t_r < 0.14$. This is approximately solved by starting at $t_r = t_{max}$ and increasing t_r in small steps of 0.001 until the first instance that $\mathrm{MBL}(t_r) \geq 10$. If $t_r \geq 0.14$, then output an all-false hair mask. Otherwise, output the optimal mask M_o by thresholding response image I_r at t_{r0}.

5.4 k-Nearest Neighbor Classification

The optimal mask M_o found from threshold search may contain some artifacts that cannot be simply removed by morphological analysis. For example, there may be spurs connected to hair pixels that originate from the microrelief. The microrelief may appear hair-like, but most would not share the same colors as hair. As such, in this section, we describe a method to take color into account for improved classification accuracy.

We use a k-NN classifier [7] to further classify hair pixels in optimal mask M_o into skin and hair pixels. Simply put, each hair pixel in $M_o$ is a query point in an m-dimensional metric space, and we find the k nearest training data points to the query point. Then, a majority rule decides how the query point is classified. For our application, the Euclidean distance between data points is used and k is set at 5. k is chosen large enough so that classification noise is minimized while k is small enough so that the structure of the 4-dimensional neighborhood surrounding the query points is preserved.

In order to use the classifier, the training data points and query points need to be defined. 4-dimensional data points of each pixel location in hair-occluded skin

image I are created by appending response I_r to the 3 channels of the skin image; we represent the data points as a data matrix $\mathbf{X}$ of size $n_{pixels} \times 4$, where n_{pixels} is the number of pixels in image I. The 3 color channels are represented in Lab color space. This color space was chosen because the Euclidean distance between data points in Lab color space is perceptually uniform. The data points in $\mathbf{X}$ are scaled to have zero-mean and standard deviation of 1.

Training data with skin pixel classification originates from the complement of optimal hair mask M_o, denoted by M_s. Pixels in $\mathbf{X}$ that are true in M_s are extracted and given the label "skin".

Training data with hair pixel classification are constructed as follows: First, an optimal hair-only threshold t_{rh} is found by increasing t_r above t_{r0} such that $\mathrm{ED}(t_{rh}) = s_r\mathrm{ED}(t_{r0})$ is satisfied, where $s_r = 0.9$. This is approximately solved by starting at $t_r = t_{r0}$ and increasing t_r in small steps of 0.001 until the first instance that $\mathrm{ED}(t_r) \leq s_r\mathrm{ED}(t_{r0})$. Second, the hair-only mask M_h is extracted from I_r using threshold t_{rh}. Third, pixels in $\mathbf{X}$ that are true in M_h are extracted and given the label "hair".

Query data comes from optimal mask M_o. Pixels in $\mathbf{X}$ that are true in M_o are extracted. By only querying these pixels, the set of hair pixels contained in the output mask from the classifier M_{kNN} is a subset of the set of hair pixels from M_o.

In Figure 5-16, M_s, M_h and M_o is shown for the test image in Figure 5-5a with response image in Figure 5-12d.

(a) (b) (c)
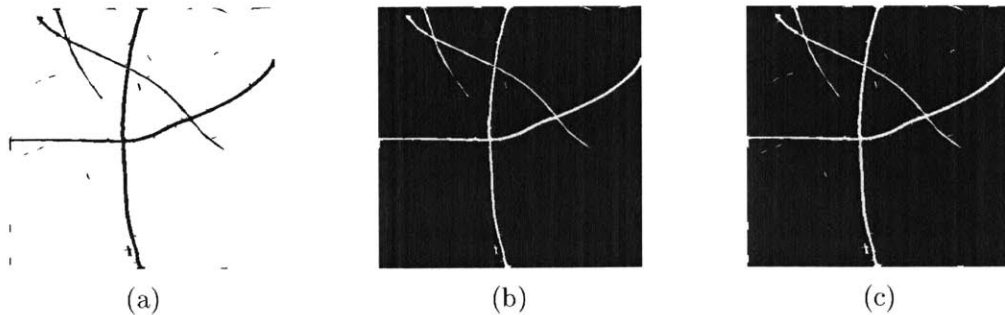
Figure 5-16: Input masks for k-NN classifier for test image with prominent microrelief (Figure 5-5a). (a) Skin-only mask M_s. (b) Hair-only mask M_h. (c) Query mask M_o.

5.5 Morphological Processing

The output hair mask from k-NN M_{kNN} contains connected regions representing non-hair objects. These are filtered out by morphological operations using 8-connectivity of pixels. First, small objects ≤ 10 pixels in area are filtered out. Second, the circularity of each connected region [104] is defined:

$$F_c = \frac{A_c}{\pi R_c^2} \tag{5.13}$$

where A_c is the area of region in pixels and R_c is the radius of the minimum bounding circle of the region. F_c takes on values of $[0, 1]$. A high F_c indicates that the region appears circular, and a low F_c indicates a hair-like region. Furthermore, objects with high F_c and large R_c may represent regions of heavily clumped hairs that should not be filtered out. Therefore, using empirically-determined values, if $F_c > 0.18$ and $R_c < 20$ for a connected region, it is filtered out as a non-hair object.

For the test image shown in Figure 5-5a, we display M_{kNN} and the final hair mask M_f in Figure 5-17.

(a)   (b)

Figure 5-17: Hair mask after k-NN classification and final hair mask for test image (Figure 5-5a). (a) Hair mask after k-NN classification M_{kNN}. (b) Final hair mask M_f after morphological operations.

5.6 Experimental Results

5.6.1 Comparison with Other Hair Segmentation Methods

Our hair segmentation algorithm was tested on the hair-occluded skin image dataset with hand-annotated ground truth shown in Figure 5-2. Accuracy is defined as follows:

$$\text{Accuracy} = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \tag{5.14}$$
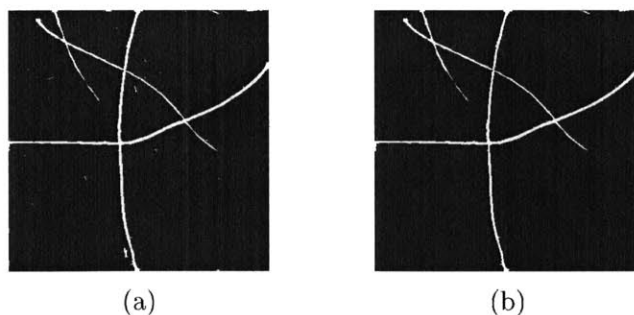
The mean sensitivity, specificity and accuracy for our method and 4 other hair detection algorithms were calculated across the 15 images (Table 5.2).

Table 5.2: Sensitivity, specificity and accuracy for hair segmentation algorithms for hair-occluded image dataset in Figure 5-2.

Method	Sensitivity	Specificity	Accuracy
Abbas et al. [3]	0.6767	0.9723	0.9370
DullRazor [53]	0.5197	0.9793	0.9486
Fiorese et al. [21]	0.9049	0.9444	0.9356
Xie et al. [104]	0.8875	0.7276	0.7260
Our Method	0.7411	0.9890	0.9591

Our method exhibited the highest specificity and accuracy among compared methods. These improvements were made due to accurate identification of non-hair objects and pixels in various stages of our algorithm. Accuracy is highly weighted towards specificity because majority of the pixels are skin pixels. However, sensitivity for our method falls short of the results for Fiorese et al. and Xie et al. In the former case, we note that hair are objects that are only a few pixels wide and a high ratio of hair pixels are adjacent to skin pixels, hence segmented hair that is just one pixel thicker can result in greatly reduced sensitivity for our proposed method. In the latter case, despite the high sensitivity, many FPs resulted in low accuracy. Among the methods with very high specificity, our method reported the best sensitivity.

We display a few representative hair mask results from various methods and the hand-annotated ground truth in Figure 5-18, which are overlaid on the original images.

134

The masks from Abbas et al. were mostly able to differentiate hairs from microrelief, but a few spurs remained on the hairs. The well-known DullRazor only performed well on thick and dark hairs. It was unable to detect thin or faint hairs. Fiorese et al. uses Otsu's method for thresholding, so there is a strong preference towards equal numbers of both classes; this led to oversegmentation when hairs were sparsely distributed. The method used by Xie et al. is very sensitive to pixels that possess similar intensity values. In cases where pixels pertaining to microrelief or melanin pigmentation on the skin and hair pixels were similar, oversegmentation occurred. On the other hand, our method accurately identified most visible hair pixels without requiring change any of its parameters to accurately segment hairs from any particular skin image. It performed reasonably well even on low-contrast and thin, faint hairs. In most cases, hair-like microrelief structure were not erroneously segmented.

5.6.2   Computation Time

The mean computation times for various hair detection algorithms are shown in Table 5.3. For images of size 400×400 from the hair-occluded skin image dataset (Figure 2-3), our algorithm was only 2 to 3 times slower than our implementations of compared algorithms in MATLAB. For images of size 1024×768 from the dermoscope dataset (Figure 5-4), a state-of-the-art algorithm by Koehoorn et al. [45] reported a mean runtime of over a minute using graphics processing unit-accelerated (GPU-accelerated) hardware. In comparison, our method was on average 5.5 times faster in MATLAB.

5.6.3   Dermoscope Images

We applied our hair segmentation method on dermoscope images with PSLs, shown in Figure 5-4. We observed that our method was able to find most hair pixels despite different illumination levels between images and varying skin background colors. However, several images contain challenging features that our method wrongly segmented as hair pixels. In images 2, 9 and 10, the skin pigmentation in the PSL appeared as
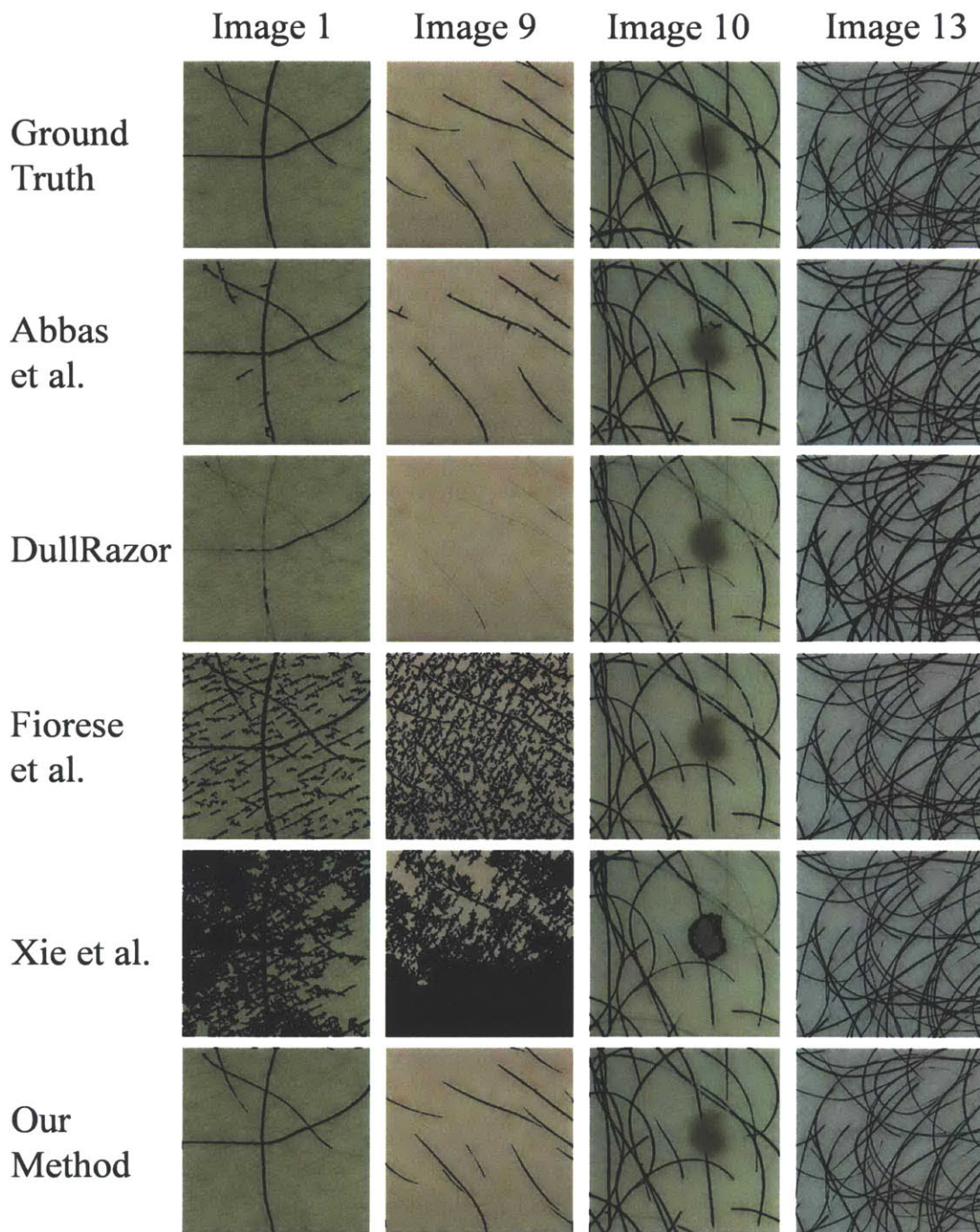
Figure 5-18: Hair masks overlaid on original images for various hair segmentation algorithms and ground truth. Pixels classified as hair are depicted in black.

Table 5.3: Mean computation times for various hair segmentation algorithms and image sizes.

Method	Runtime (s)	Implementation
Hair-Occluded Skin Image Dataset: 400 × 400 RGB Images		
Abbas et al. [3]	0.8523	MATLAB
DullRazor [53]	~0.4	C
Fiorese et al. [21]	1.2633	MATLAB
Xie et al. [104]	1.2572	MATLAB
Our Method	2.5003	MATLAB
Dermoscope Dataset: 1024 × 768 RGB Images		
Koehoorn et al. [45]	73.8	GPU-accelerated C++
Our Method	13.4	MATLAB

short, elongated structures with similar colors to hair, hence these structures were segmented erroneously. Image 4 was a particularly challenging case as small circular holes on the surface of the lesion were classified as hair even though they should be filtered out by morphological processing. Moreover, image 5 indicated that our algorithm wrongly detected lines from air bubbles in the dermoscope immersion fluid as hair. Lastly, we noticed that in image 9, our method was able accurately segment a fair amount of clumped hair.

5.6.4 Light Hair Detection on Dark Hairs

In our proposed method, the black top-hat transform is substituted with the white top-hat transform in order to detect light hairs. Ideally, for images with dark hairs only, no pixels should be marked as light hairs. When we ran our algorithm on the hair-occluded skin image dataset (Figure 5-2), no light hair pixels were detected for most images. However, images 4 and 10 contained some pixels labeled as light hair, as seen in Figure 5-20. These "light hairs" were actually thin, elongated gaps between dark hairs. As a direction for future work, we recommend morphological analysis of the dark hair mask to determine the likelihood that pixels in the light hair mask are true light hairs.

137

Image 1

Image 2

Image 3

Image 4

Image 5

Image 6

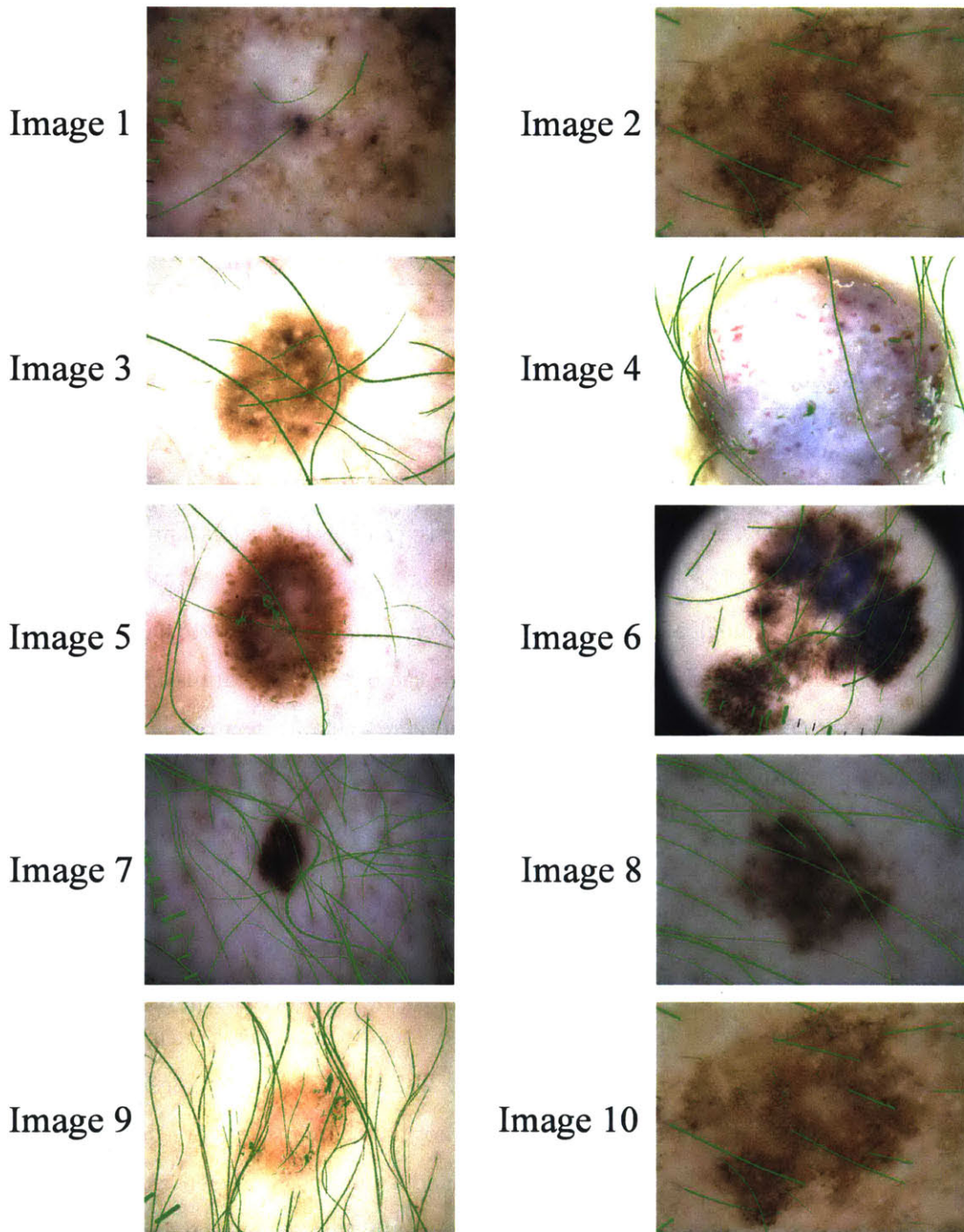Image 7

Image 8

Image 9

Image 10

Figure 5-19: Hair masks from our method overlaid on dermoscope images containing PSLs and dark hairs. Pixels classified as hair are depicted in green.
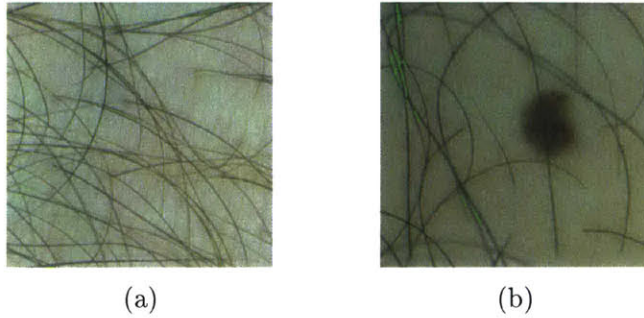
<center>(a)   (b)</center>

Figure 5-20: Hair masks found using light hair detection on images with only dark hairs. Green pixels in images indicate pixels classified as hair. (a) Image 4. (b) Image 10.

5.6.5 Low-Contrast Hairs

The image of a mixture of dark and light hairs (Figure 5-5b) presents challenging circumstances for our hair segmentation algorithm. Both dark and light hairs in the image are thin, low-contrast features. Although our multiscale matched filter is able to highlight the thin, single-pixel thick light and dark hair structures (Figure 5-10), the adaptive threshold declared failure by returning an all-false hair mask in both cases. In the plots of $ED(t_r)$ and $MBL(t_r)$ for dark and light hairs in Figure 5-21, MBL never exceeded 10 and peak ED was located at relatively high values of t_r. Thus, our adaptive threshold failed to detect hairs. For these two cases, the optimal threshold was actually above $t_r = 0.14$, which is difficult to predict based on ED and MBL measures alone. Therefore, a direction for future work is that of improving the adaptive threshold by finding new mask measurement functions so as to detect low-contrast hairs.

5.7 Discussion

In this chapter, we presented an automatic hair segmentation algorithm that achieved very accurate segmentation results. It combines the top-hat transform and multiscale matched filter to detect hairs. Next, it is employs a robust adaptive threshold. The thresholded result is refined by k-NN classification of hair and skin pixels and morphological processing. Compared to other hair segmentation algorithms in our evaluation,

<center>139</center>
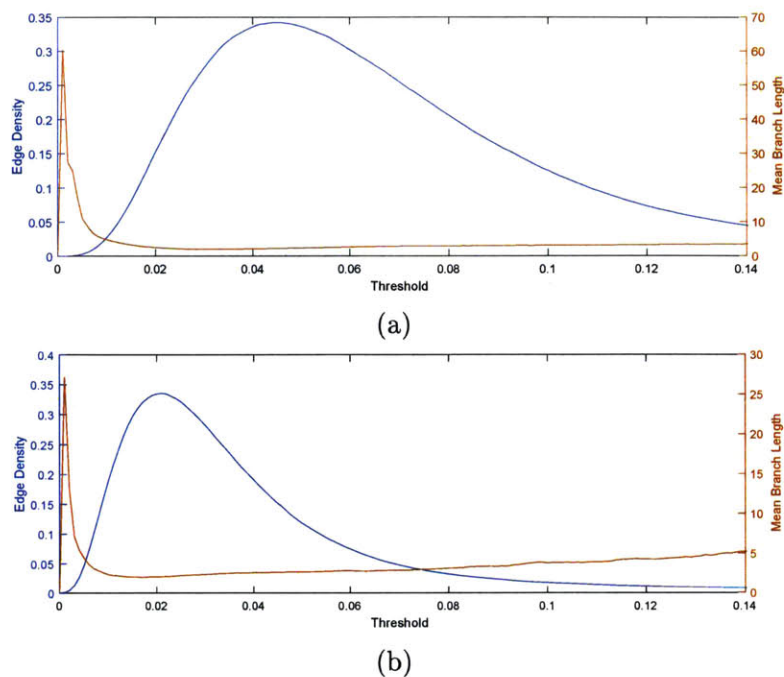
Figure 5-21: Graphs of ED and MBL against threshold values for image with mixture of dark and light hairs. (a) Using dark hair detection. (b) Using light hair detection.

our proposed algorithm performed much better in producing visually coherent hair masks for skin images with a variety of dark hairs in a range of colors, hair distributions and widths: most non-hair structures such as microrelief were removed and hairs extracted were smooth, elongated lines with accurate widths. Thus, our method is more robust than existing hair segmentation algorithms. At this level of accuracy, the algorithm is fast; it is 5 times faster than a recent GPU-accelerated method. In the same vein as existing literature, we tested our method on a set of dermoscope images with PSLs. Our algorithm detected almost all dark hairs and comparatively few non-hair objects remained in the mask.

Nevertheless, our proposed method has limitations that remain challenging research problems in the field. First, combining detection of dark and light hairs remains problematic. As dense dark hairs produce erroneous light hairs between thin gaps, a robust technique to combat this problem is desirable. Second, although our curvilinear structure detector was able to detect low-contrast dark and light hairs, the adaptive threshold was unable to decide an appropriate threshold level. Hence,

a more robust thresholding method based on new mask measurement functions will be explored in the future. Third, in our datasets, we used images with even illumination. In more realistic settings, local illumination varies within an image, and different adaptive thresholds may be needed in different regions of the same image. Thus, as future work, we will explore different methods to vary the threshold across the spatial domain of a skin image. Examples of these methods include fusing information on local illumination levels into our adaptive threshold or initially dividing the image into separate visually coherent connected regions. Fourth, several skin images contain low-contrast, faint hairs with very large gaps and hairs that were out of focus. These hairs were not reliably detected by our curvilinear structure detection method. To increase accuracy, predictive tracking of individual hairs is desired. This research direction also paves the way for hair counting and accurate width determination, which have various applications areas such as tracking the progression of hair loss.

For predictive hair tracking, we propose using the multiscale matched filter to reliably detect the orientation of individual hairs based on the orientation angle that gives the maximum response at each pixel. An illustration of the result is shown in Figure 5-22b for the image shown in Figure 5-22a. The orientation at each hair pixel in the image is similar to the orientations of its neighboring hair pixels, even for pixels in very low-contrast hairs regions, so connected regions with similar orientations can be easily extracted. This allows the full width of hairs to be approximated. Assuming that the orientation of pixels of hairs in connected regions with similar orientations do not abruptly change and that hair width remains approximately constant, tracking by orientation is done as follows: Using seed regions derived from our original proposed method, individual hairs are identified and tracked by traversing the length of successive hair regions with similar orientations while keeping the hair widths approximately constant to ensure smoothness of hair edges. If hair gaps appear, the lengths of the ends of hair can be extended based on similar orientations of two ends of hair that are nearby while maintaining an appropriate hair width.
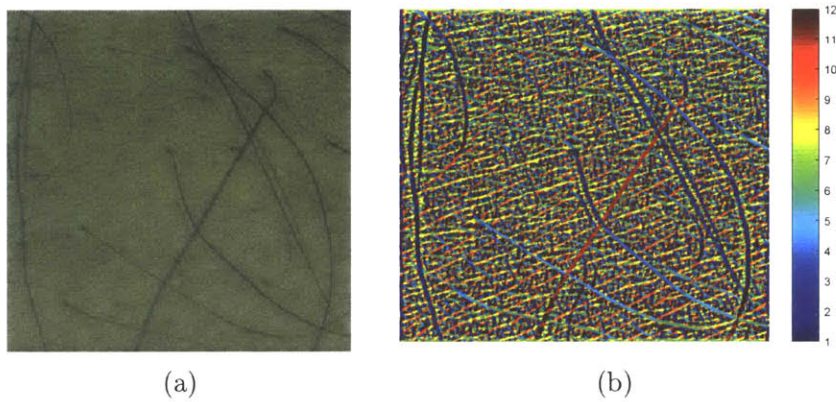
Figure 5-22: Hair orientations at each pixel from multiscale matched filter and corresponding skin image. (a) Skin image with hairs. (b) Orientation of hairs at each pixel. The color bar represents orientations in steps of $\theta = \{0°, 15°, \ldots, 165°\}$ mapped to orientation numbers $\{1, \ldots, 12\}$.

Chapter 6

Guided Inpainting of Skin Images

Tying together our skin registration and hair segmentation algorithms, we present a novel method for reconstruction of the hair-occluded skin surface. Using multiple viewpoints of the skin surface and physically shifting hairs at the same time (i.e. combing hairs in different directions), registered skin images with digitally removed hair are combined to construct actual skin texture. This is motivated by the need to remove hairs from skin images to reveal the underlying microrelief structure for skin image analysis, particular those with hairs that are dark, long, thick and dense. In Chapter 2.4.1, we described that inpainting methods currently used in digital hair removal (DHR) methods simply interpolate intensity values from skin pixels that border hair pixels. They contain information on true skin texture obstructed from view by the hairs. In our evaluation, we will visually compare these conventional inpainting methods against our guided inpainting method.

6.1 Algorithm Design

6.1.1 Overall Design

The overall design of guided inpainting for skin images is illustrated in Figure 6-1. The goal is to reconstruct skin texture obstructed by hairs in target image T. First, target image T and N source images $S_1, \ldots, S_N$ are each processed by the

hair segmentation method described in Chapter 5. The target and source images are multiple viewpoints of the same skin patch scene with hairs physically moved before each image is acquired. Then, hair binary masks from segmentation are dilated by a square structuring element of 5 pixels in width. This step ensures that the entire hair border of detected hairs is delineated. Otherwise, darkened hair borders may appear on the final reconstructed image. This is followed by initial inpainting using MATLAB proprietary function *regionfill* for all images. The function smoothly interpolates intensity values of inner hair pixels from skin pixels at the borders by solving Laplace's equation for Dirichlet boundary conditions. Initial inpainting prevents discontinuities at the boundaries of removed hair so that image patches can be more reliably registered. Also, implicitly, initial inpainting is used to smoothly "fill-in" hair pixels in the target image that remain unknown due to occluding hair or inaccurate registration in all source images. At the end, hair masks (denoted by M_T and $M_{S_1}, \ldots, M_{S_N}$) and initially inpainted skin images (denoted by T' and $S'_1, \ldots, S'_N$) are obtained. Second, initially inpainted source images $S'_1, \ldots, S'_N$ are each registered to initially inpainted target image T' using our skin registration method described in Chapter 3. Thereafter, registered images $R_1, \ldots, R_N$ are acquired. Third, missing pixels values in initially inpainted image T' are replaced according to information in registered images $R_1, \ldots, R_N$ which we term *guided inpainting*. Guided inpainting consists of two steps: (1) compositing an (2) blending, which will be described in the following sections. The output image of guided inpainting is denoted T_I.

6.1.2 Guided Inpainting

Compositing

Guided inpainting is motivated by discontinuities in intensity at the boundaries of hair and skin as a result of compositing. These discontinuities are seen as visible seams. In the usual sense, compositing is the direct replacement of missing pixel values in the target image from the registered image. In our case, for N source images, we define compositing as directly replacing missing hair pixel RGB values in intially inpainted
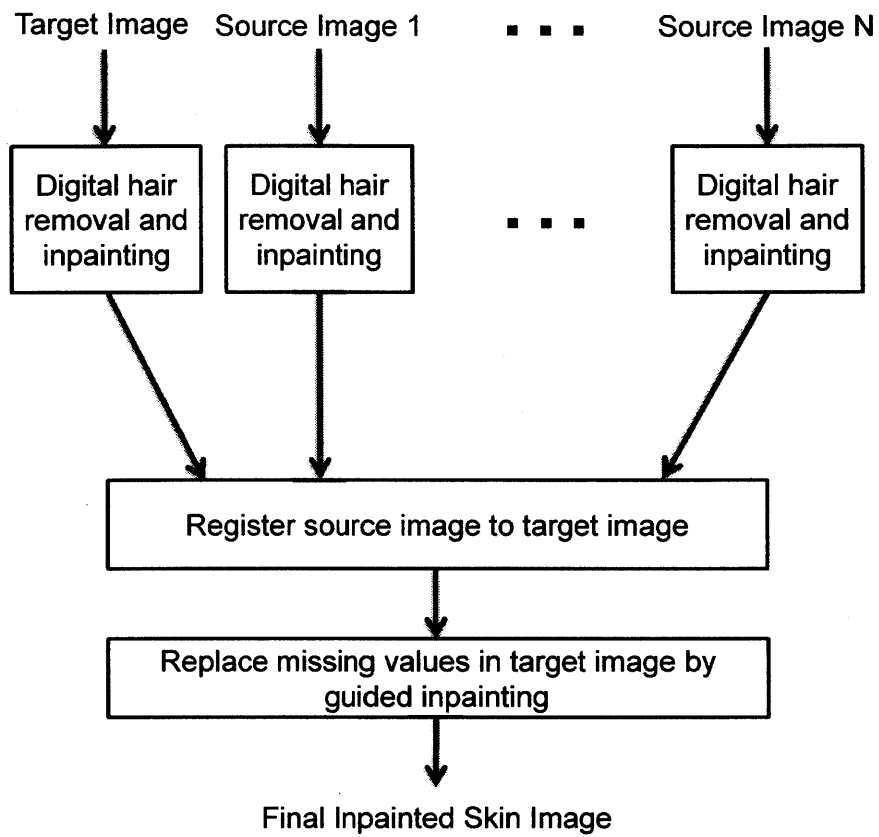
144

Figure 6-1: Algorithm flowchart for overall design of guided inpainting of skin images.

image T' with values chosen from skin pixel values in registered images $R_1, \ldots, R_N$. Given a pixel classified as hair in image T', the compositing procedure described below assigns each hair pixel to an image in registered image set $\{R_i \mid i = 1, \ldots, N\}$ such that the corresponding skin pixel in R_i replaces the hair pixel.

First, using the dense deformation field estimates from skin registration, hair masks of source images $M_{S_1}, \ldots, M_{S_N}$ are registered. Registered hair masks are denoted by $M_{R_1}, \ldots, M_{R_N}$.

Second, for each image pair T' and R_i, the spatial distribution of final matched keypoints, denoted by K_f^i, from skin registration determines which regions of the corresponding registered image are aligned with the target image. For simplification of notation, as K_f^i actually represents a pair of keypoints, we let K_f^i refer only to keypoints in the domain of initially inpainted target image T'. In general, a high local density of K_f^i indicates a high likelihood that the local region is accurately registered. As such, keypoints in K_f^i are removed unless there are at least $K_{fn} = 3$ keypoints found within a radius $K_{fr} = 100$ from that keypoint. K_{fn} and K_{fr} are chosen such that most isolated keypoints are eliminated so that the filtered set of keypoints maintains a sufficiently dense and even spatial distribution of keypoints. Large values of K_{fn} and small values of K_{fr} tend to filter out keypoints in sparsely populated regions of keypoints, while small values of K_{fn} and large values of K_{fr} tend to retain more isolated keypoints. The filtered set of keypoints is denoted by K_{ff}^i.

Third, we let each keypoint in K_{ff}^i determine an enclosed region indicated by a filled circle of radius v_r with center coordinates being the coordinates of the keypoint. v_r is equal to the maximum distance of neighbors found during keypoint removal, or equivalently, the distance to the K_{fn}-th nearest neighbor in K_f^i. For an image pair T' and R_i, the union of all filled circles is termed the *region of validity*. This is indicated by true values in binary mask M_v^i.

Fourth, a binary mask M_{vh} is constructed to indicate hair pixels in initially inpainted image T' with known skin pixel values due to the registered images. The mask M_{vh} is computed by taking the intersection of target hair mask M_T with the union of all skin masks from registered source images (denoted by $\overline{M}_{R_1}, \ldots, \overline{M}_{R_N}$) as

follows:

$$M_{vh} = M_T \cap (\overline{M}_{R_1} \cup \ldots \cup \overline{M}_{R_N}) \tag{6.1}$$

Fifth, M_{vh} is constrained to the region of validity by computing the intersection of M_{vh} with the union of all regions of validity from the N target and registered image pairs:

$$M'_{vh} = M_{vh} \cap (M_v^1 \cup \ldots \cup M_v^N) \tag{6.2}$$

This ensures that the resulting binary mask M'_{vh} contains only hair pixels in initially inpainted image T' with known skin pixel values from accurately registered regions of the registered source images.

Sixth, label assignments from $1, \ldots, N$ are decided by partitioning pixel locations in regions marked true in mask M'_{vh}. For each true pixel in mask M'_{vh}, the distance to the nearest keypoint in each set of filtered keypoints K^i_{ff} is calculated. Then, label assignments from $1, \ldots, N$ for each pixel in mask M'_{vh} is given by the minimum distance among nearest keypoints. Sixth, using the label assignment, hair pixels in T' are replaced by skin pixels from corresponding registered images $R_1, \ldots, R_N$. The composite image is denoted by C.

In Figure 6-2, we display the algorithm flowchart of the compositing process using notation defined in the above description.

Image Blending

Initially inpainted target image T' and composite image C are blended using Laplacian pyramid blending [12] with mask M'_{vh}. In this process, image C *guides* initially inpainted target image T' such that skin textures are preserved in the final inpainted skin image, denoted by T_I. Note that if a pixel in C is composited, then skin texture is preserved at that pixel, otherwise, blending ensures intensity values at all other regions are close to preserved, including those of initially inpainted regions in T'.
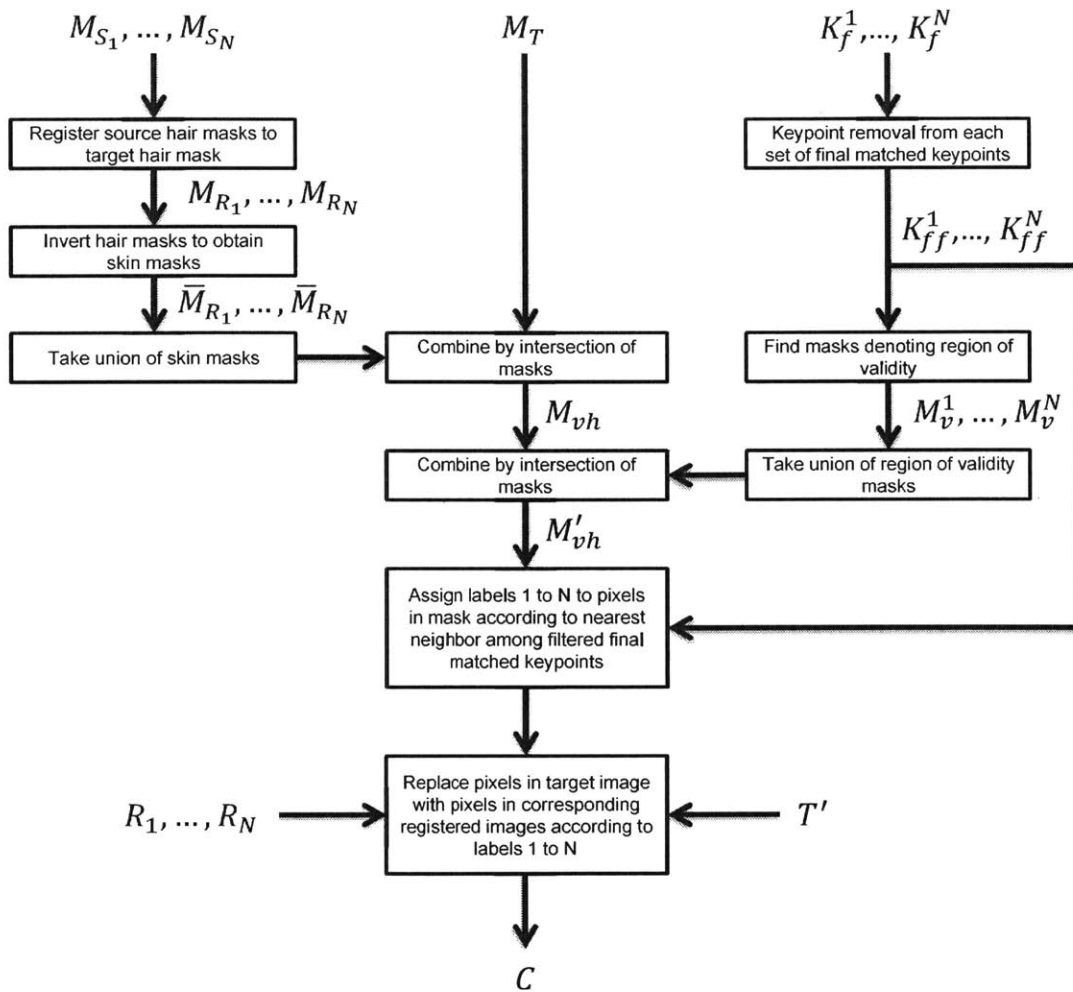
147

Figure 6-2: Algorithm flowchart for compositing. Notation follows description found in Chapter 6.1.2.

To generate a Gaussian pyramid, an image is successively blurred by a Gaussian filter and downsampled. This generates a set of low-pass filtered images with decreasing filter bandwidths; each image represents one of the l_p levels in the pyramid. The Laplacian pyramid is a set a band-pass filtered images constructed by taking the difference between consecutive images in Gaussian pyramid levels. A detailed exposition on these pyramids, including decomposition, reconstruction and choice of parameters, is found in [12].

To blend both images, the following steps are taken:

1. Build Laplacian pyramids $L_{T'}$ and L_C for images T' and C respectively.

2. Build a Gaussian pyramid $G_{M'_{vh}}$ for mask M'_{vh}.

3. For each level l_p and pixel location (x, y), combine $L_{T'}$ and L_C using weights from $G_{M'_{vh}}$ to form a combined pyramid L_{T_i}:

$$L_{T_i}(l_p, x, y) = G_{M'_{vh}}(l_p, x, y)L_C(l_p, x, y) + \left(1 - G_{M'_{vh}}(l_p, x, y)\right)L_{T'}(l_p, x, y)$$

4. Collapse pyramid L_{T_i} by successively upsampling and summing each level to produce final image T_I.

As inpainted regions in image T' are smooth, there are few high frequency details in these regions. Hence, the overall effect is that high frequency details from composited pixels in image C are added to inpainted regions in T', while the low frequency content (i.e. color) in these regions in C and T' are blended to remove seams.

6.2 Experimental Results

6.2.1 Implementation

Our procedure was implemented in MATLAB Release 2015b. A system with an Intel Xeon E3-1225 v3 3.2 GHz processor was used. Also, we used the OpenCV 2.4.13

(Python 2.7.12) [8] implementation of the fast marching method [93] and Windows application DullRazor [53] for comparison of different inpainting methods.

6.2.2 Dataset

The aim of this investigation is to provide a proof-of-concept for guided inpainting of skin images. Therefore, a set of 4 hair-occluded skin images, comprising 1 target image and $N = 3$ source images, were obtained from the same area of the ventral forearm. The images are of resolution 2040×2040. The images were taken from a male of Caucasian descent in the 18 to 35 age group. The images exhibit dark, long, thick and dense hairs on tanned skin. A human operator set the camera orientation constant while the long hairs were physically moved by combing hairs in different directions before recording each image to reveal different regions of the underlying skin. The set of 4 images are shown in Figure 6-3. The red square outlines inked on the skin surface only serve as visual markers for human observers to validate registration results; they are not used by the skin registration algorithm for alignment.

(a)　　　　　(b)　　　　　(c)　　　　　(d)
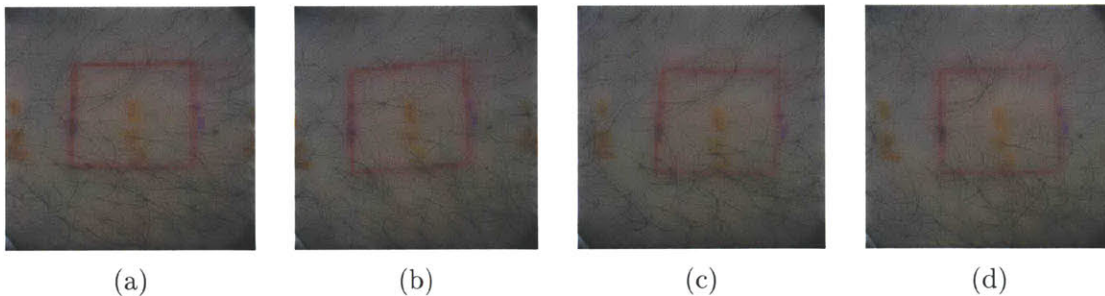
Figure 6-3: Dataset for guided inpainting of skin images. (a) Target Image T. (b) Source image S_1. (c) Source image S_2. (d) Source image S_3.

6.2.3 Algorithm Demonstration

Corresponding hair masks after hair segmentation for our dataset is displayed in Figure 6-4. After initial inpainting and subsequent image registration, we obtained initially inpainted target image T' and registered images R_1, R_2 and R_3, as shown in Figure 6-5. Most hair pixels in the images were removed but some degree of

150

misalignment occurred in different local regions. Sets of final matched keypoints from $K_f^1, \ldots, K_f^3$ found during registration are shown in Figure 6-6. Green keypoints represent keypoints remaining after keypoint removal $K_{ff}^1, \ldots, K_{ff}^3$, and red keypoints represent keypoints removed. The locations of matched keypoints are correlated with locally aligned regions of the image. This is reflected in Figure 6-7, which shows the respective regions in binary masks M_v^1, M_v^2 and M_v^3 where true pixels in the masks represent regions that are deemed to be accurately registered. Figure 6-8 displays the assignment of pixels in M_{vh}'. Blue pixel assignments indicate label 1 and correspond to pixels taken from R_1 to create composite image C. Similarly, red pixel assignments indicate label 2 and green pixel assignments indicate pixel label 3. Black pixel assignments indicate pixels that are taken from image T' (i.e. pixels are not replaced).

| (a) | (b) | (c) | (d) |
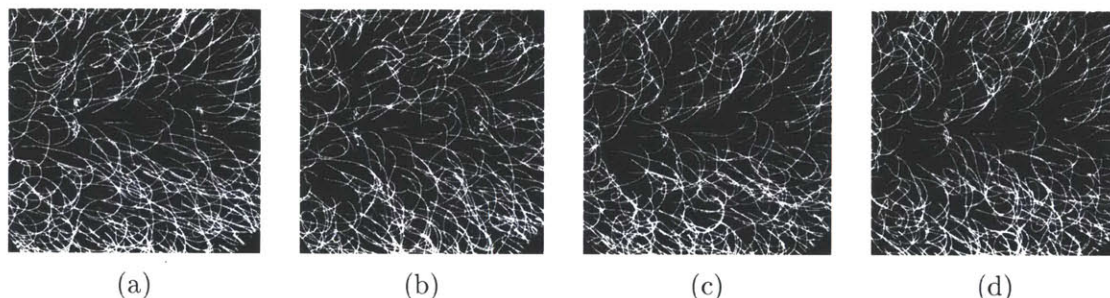
Figure 6-4: Hair segmentation masks from target and source images. (a) Target image mask M_T. (b) Source image mask M_{S_1}. (c) Source image mask M_{S_2}. (d) Source image mask M_{S_3}.

| (a) | (b) | (c) | (d) |
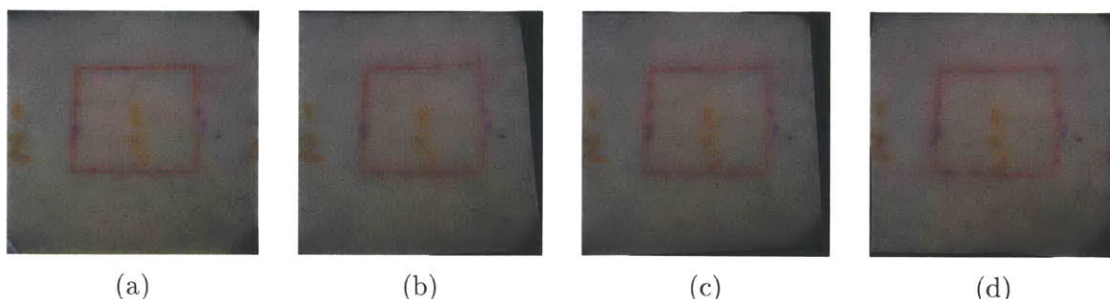
Figure 6-5: Target and source images after initial inpainting and image registration. (a) Initially inpainted target image T'. (b) Registered image R_1. (c) Registered image R_2. (d) Registered image R_3.
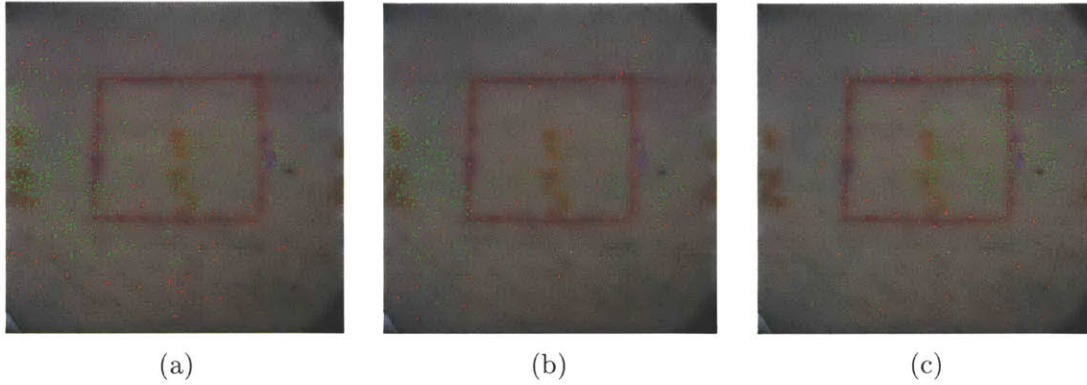
Figure 6-6: Sets of final matched keypoints $K_f^1, \ldots, K_f^N$, found during image registration and after keypoint removal. Green keypoints represent keypoints remaining after keypoint removal $K_{ff}^1, \ldots, K_{ff}^N$. Red keypoints represent keypoints removed by keypoint removal. (a) Image pair T' and R_1. (b) Image pair T' and R_2. (c) Image pair T' and R_3.
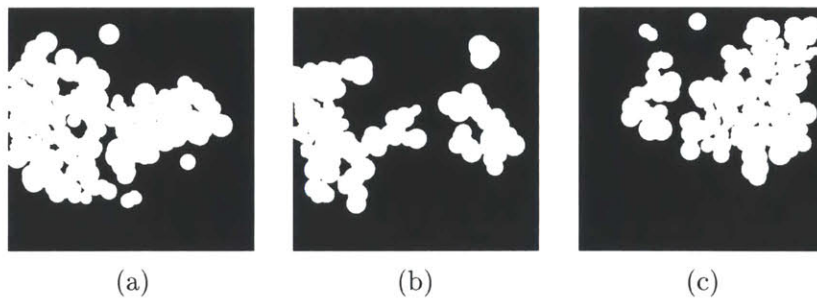
Figure 6-7: Region of validity in different image pairs indicated by true pixels in binary masks. (a) M_v^1. (b) M_v^2. (c) M_v^3.
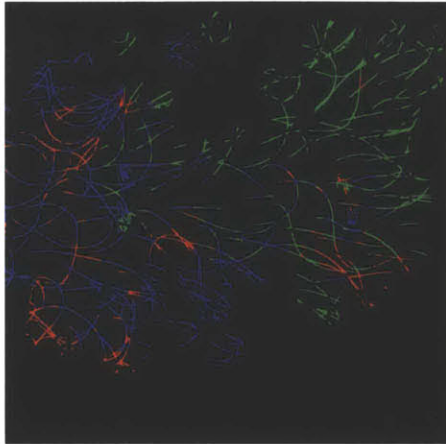
Figure 6-8: Assignment of pixels in mask M'_{vh}. Blue pixels indicate label 1, corresponding to pixels taken from image R_1. Red pixels indicate label 2, corresponding to pixels taken from image R_2. Green pixels indicate label 3, corresponding to pixels taken from image R_3. Black pixels indicates null assignment; pixels are not replaced in T'.

6.2.4 Visual Comparison with Other Methods

By visual inspection, we compared final inpainted images T_I from our guided inpainting method and various other methods. The full-sized images can be seen in Figure 6-9. A region of size 300×300 is extracted from all full-sized images at the top-left corner of the red square visual marker on the skin. These images are shown in Figure 6-10. The original image in (a) shows dense, thick hairs occluding the skin surface. The average image (b) and median image (c) of images initially inpainted image T' and registered images R_1, R_2 and R_3 drastically reduced the appearance of remaining hair objects in the image, but the microrelief was significantly blurred as the images were not all aligned accurately throughout. For (d), DullRazor [53] failed to remove many hair pixels from target image T. Wherever hair pixels were detected, bilinear interpolation caused these pixels to lose their texture and become smooth in T_I. The results in (e) and (f) appear similar to each other. Using our hair segmentation algorithm on target image T, most dark hair pixels were removed. However, the fast marching method [93] in (e) and MATLAB's *regionfill* in (f) cannot replicate skin texture. Thus, skin texture occluded by hair appeared smooth. In (g), replacing missing pixels by our compositing procedure allowed texture to be reconstructed.

153

But seams appeared at the borders of hair objects, especially when different local illumination levels were present in the images. By blending the composite image C with the initially inpainted target image T in (h), the seams were much less obvious; reconstructed skin texture by guided inpainting looked realistic in skin areas that were composited in C.

6.2.5   Visual Comparison Using Microrelief Extraction

Accurate reconstruction of skin texture allows us to extract the microrelief structure of skin images obscured by dense hairs with improved accuracy. In this section, we ran our microrelief extraction algorithm (described in Chapter 4.1.4) on the original target image T, image inpainted by fast marching method [93] and our hair segmentation result and image inpainted by our proposed guided inpainting method. The results are shown in Figure 6-11. The original image in (a) and its corresponding 300×300 region in (d) showed that hair caused the microrelief detector to extract both the skeleton images of hair and microrelief. Also, many large gaps in the structure were present. For the image inpainted by fast marching method in (b) and its corresponding zoomed-in version in (e), we found that DHR drastically reduced the amount of gaps in the extracted structure. Also, interfering hair structures were no longer extracted. However, some gaps in the structure continued to appear at smooth inpainted skin pixels. The extracted microrelief for our proposed method, displayed in (c), with its corresponding zoomed-in version (f), showed that most gaps in the structure had disappeared. Furthermore, the microrelief was visibly more connected at inpainted regions. In this demonstration, by drastically reducing the amount of gaps in the microrelief structure, guided inpainting seems to improve the accuracy of microrelief extraction on skin images with dense hairs. As future work, we propose ground truth comparison by comparing microrelief structure extracted before shaving and after shaving hairs from the same area the skin.
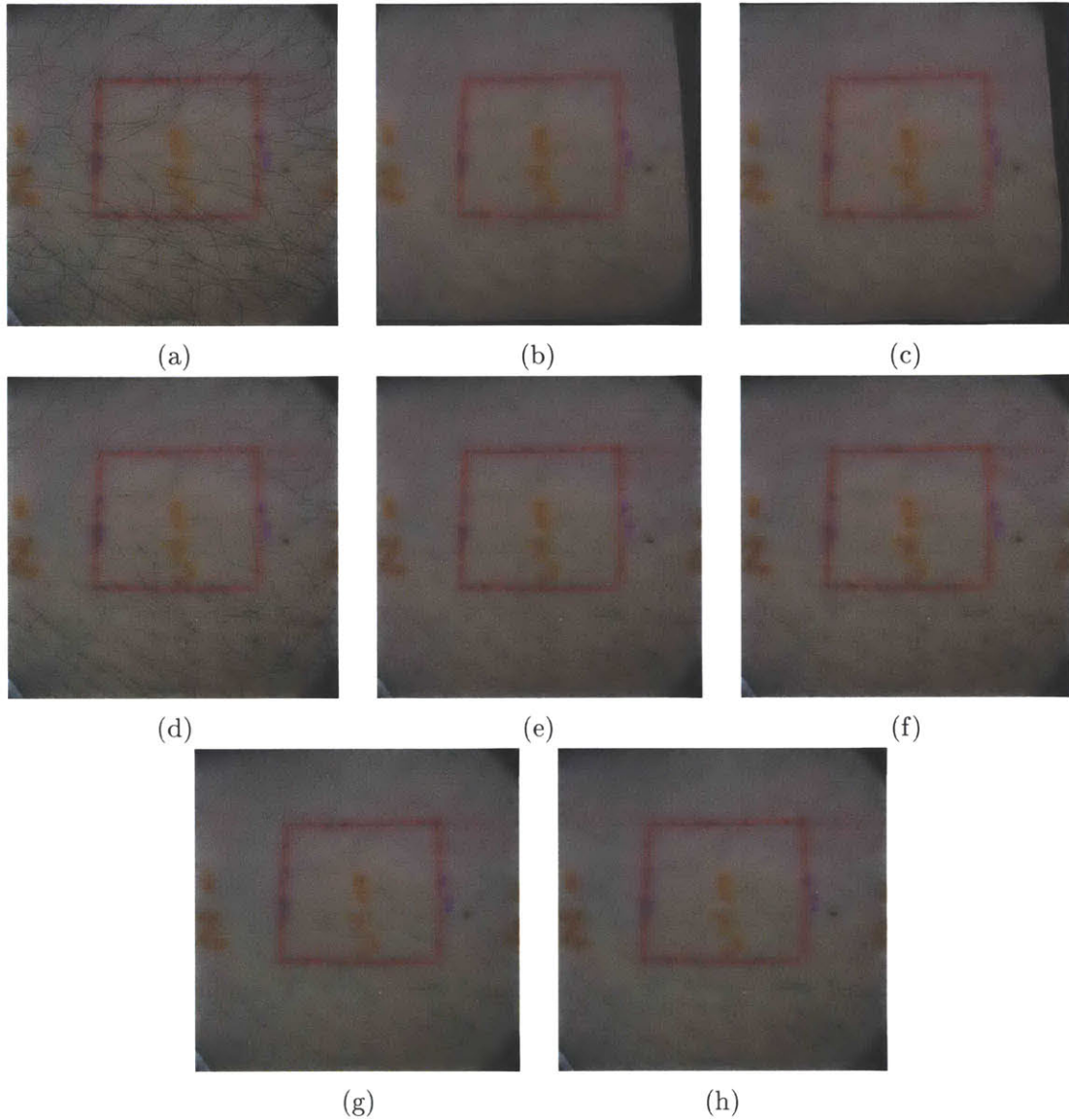
Figure 6-9: Comparison of final images T_I for various inpainting methods. (a) Original image T. (b) Average image. (c) Median image. (d) DullRazor on T. (e) Fast marching method on T. (f) MATLAB's *regionfill* on T. (g) Compositing. (h) Guided inpainting.

155
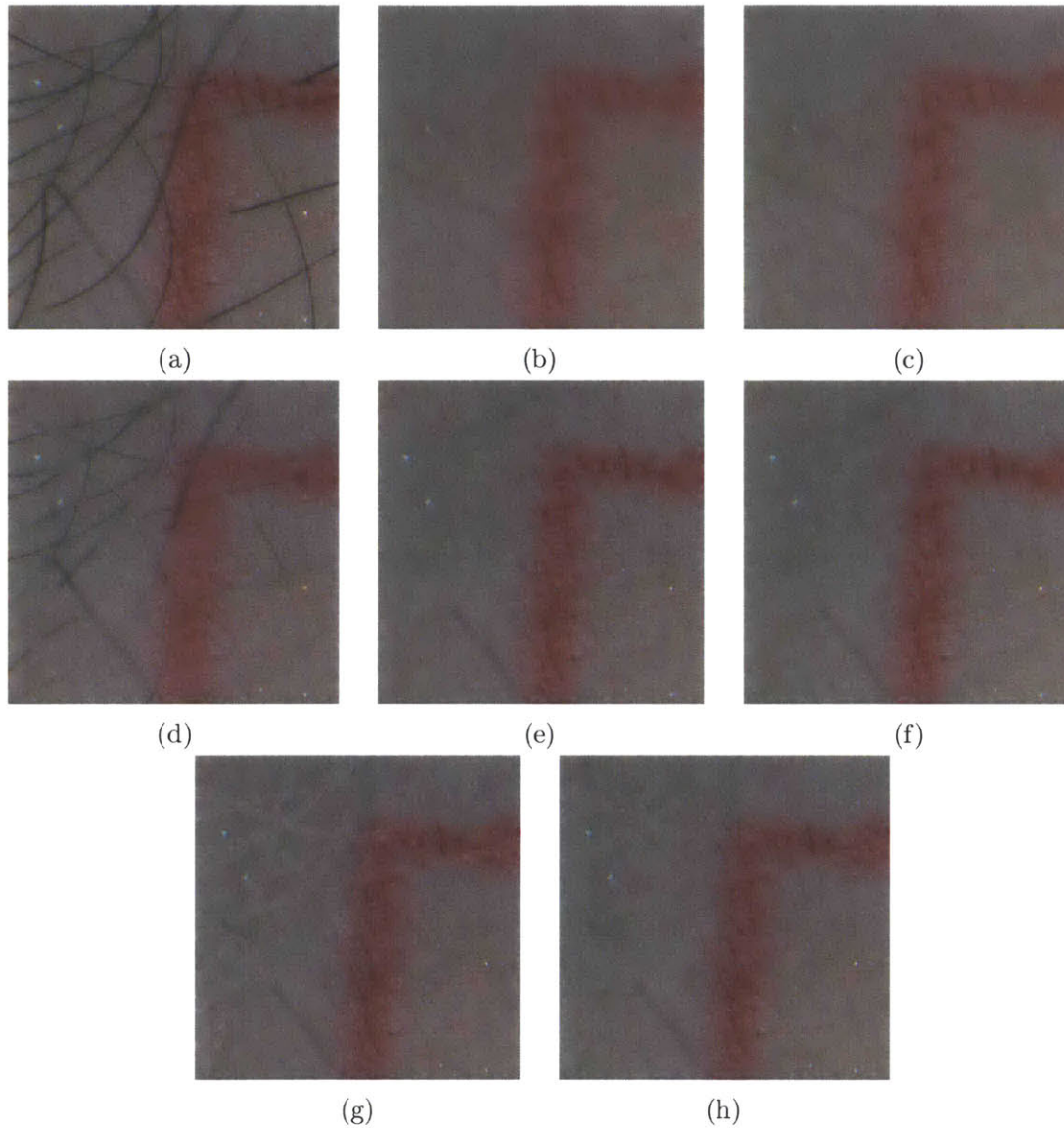
Figure 6-10: Zoomed-in comparison of final images T_I for various inpainting methods. A 300×300 region from images in Figure 6-9 is shown. (a) Original image T. (b) Average image. (c) Median image. (d) DullRazor on T. (e) Fast marching method on T. (f) MATLAB's *regionfill* on T. (g) Compositing. (h) Guided inpainting.
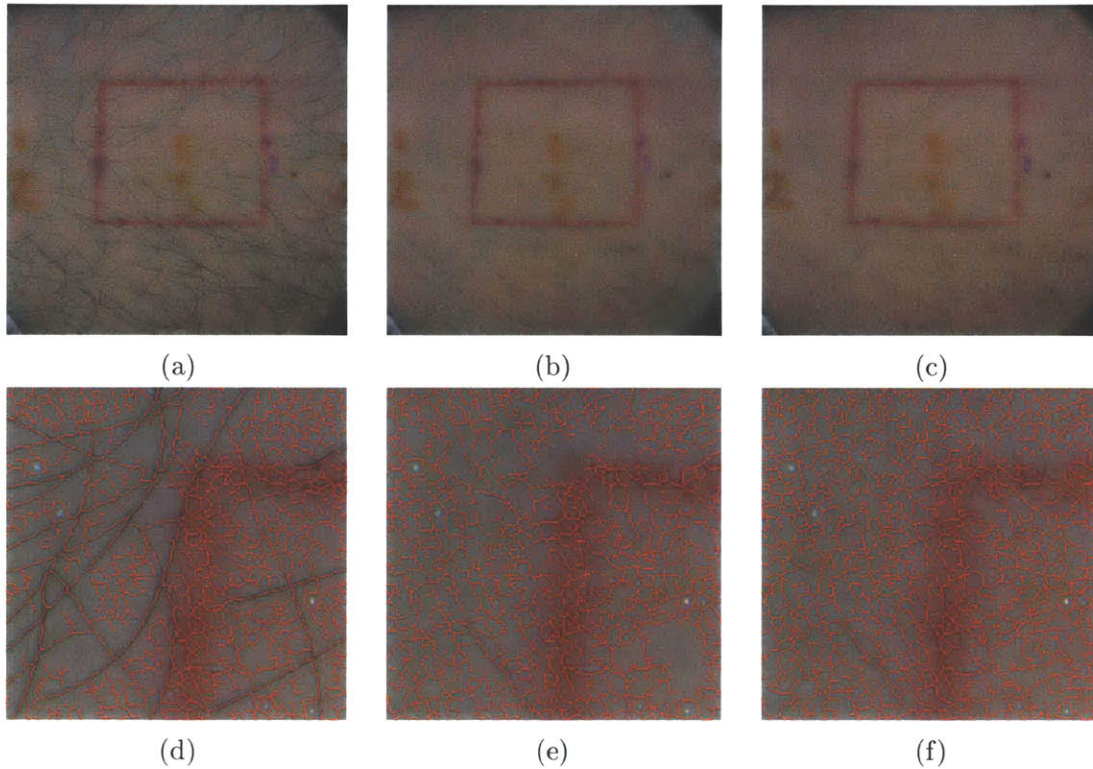
Figure 6-11: (a-c) Extracting microrelief from images with dense hair. (a) Original image T. (b) Inpainting of T by fast marching method using our hair segmentation result. (c) Guided inpainting of T by compositing and blending using our hair segmentation result. (d-f) 300×300 region of full-sized images showing extracted microrelief in detail. (d) Original T. (e) Fast marching method. (f) Guided inpainting.

6.3 Discussion

In this chapter, we described guided inpainting for reconstruction of texture on hair-occluded skin. For accurate texture reconstruction, guided inpainting uses images from multiple viewpoints of the skin with hairs physically moved around before recording the image. In a proof-of-concept demonstration, we were able to register all other images to the first image after hair segmentation and initial inpainting, then selectively composite pixels and blend images to produce a realistic-looking skin patch without hairs. Unlike conventional inpainting methods, the true skin texture is reconstructed. From this skin patch, we could extract the obscured microrelief structure of the skin more accurately due to the drastic reduction in gaps on the extracted structure compared to conventional inpainting.

There are some limitations to using this method. First, accuracy of reconstruction is dependent on accuracy of registration as a registration result that has error of only 1 or 2 pixels produces skin textures with discontinuities in the thin microrelief structure. Because the accuracy of our registration method is dependent on density of matched keypoints in local areas, only areas with sufficient density of matched keypoints can be accurately reconstructed. Poor hair segmentation and changes in skin appearance due to changes in viewpoint both negatively affect the density of matched keypoints. If the density of matched keypoints is not high enough, our method reverts to conventional inpainting (using intensity values from the initially inpainted target image). Hence, guided inpainting motivates the development of more robust and accurate hair segmentation algorithms and an imaging system that allows skin to maintain its appearance from different viewpoints.

Second, our compositing procedure currently assigns a hair pixel in the initiallly inpainting target image to a skin pixel in the image containing the nearest keypoint to the hair pixel. It may be helpful to enforce spatial regularization of assignment in order to prevent many discontinuities in pixel intensity in the composite image due to misalignment or different illumination levels of registered images. Another direction is that of making small spatial shifts in composited regions in order to

reduce misalignment of microrelief structure.

As future work, we plan to develop a qualitative evaluation procedure for measuring the fidelity of inpainted skin texture to actual skin texture, which would likely entail comparison of skin images or microrelief structure for images acquired before and after shaving hairs. Also, experiments will be conducted on a larger set of skin images so as to test guided inpainting for robustness to various skin and hair types.

Chapter 7

Conclusion

In this thesis, a set of computational tools that enable quantitative analysis of skin images taken over a period of time was presented. The enabling tools that were described and evaluated are nonrigid skin image registration, digital hair segmentation and guided inpainting of hairs. Also, a microrelief junction point matching (MJPM) algorithm was proposed as a potential way to study the stability of skin microrelief structure over time and serve as a refinement for registration.

The proposed nonrigid skin registration algorithm using multiple plane detection for landmark identification and multilevel B-splines as the local motion model was shown to register pairs of skin images within a 1 inch by 1 inch patch of skin with very low sub-pixel root-mean-square error (RMSE). Experiments with skin images containing sparse amounts of hair revealed that it is robust to transformations such as in-plane image rotations and both local and global nonlinear distortions at different scales. It is also robust to relatively high levels of camera sensor noise. Visual evaluation revealed that the method is able to accurately register skin images taken from different viewpoints, with induced stretching and taken over in the timescale of a few months, even with variations in camera settings and illumination levels. Future work for this algorithm includes the exploration of more sophisticated multiple plane detection methods for robust keypoint matching, and faster and more reliable rigid registration methods for the initial guess of the transformation.

A new digital hair segmentation algorithm was developed and it proved to pro-

vide accurate hair segmentation for a varied set of skin images containing both dark hairs and hair-like microrelief structure. Against several hair segmentation methods in published in the literature, this algorithm exhibited the highest accuracy for an dataset of 15 skin images. This was achieved with combination of high sensitivity and specificity levels. Visual inspection of the results showed that the algorithm accurately extracted the smooth contours of hair with little or no interference from non-hair objects. Testing the algorithm on dermoscope images containing pigmented skin lesions (PSL) revealed that it is able to extract most visible dark hairs accurately without erroneously extracting the diverse assortment of non-hair structures. Also, this algorithm is at least 5 times faster compared to a recent algorithm with reportedly high levels of accuracy and robustness. Nevertheless, additional research in adaptive thresholding for low-contrast hairs and merging light and dark hair detection are required in order to make this automatic segmentation algorithm more robust to different scenarios.

An approach for guided inpainting of hairs merged skin registration and digital hair segmentation into a novel method for preserving high-resolution skin details and texture. A proof-of-concept for guided inpainting demonstrated that reconstructed skin textures and microrelief occluded by hairs looked realistic compared to conventional inpainting methods. Furthermore, running microrelief extraction on the reconstructed skin image resulted in an extracted microrelief structure with fewer gaps compared to that of a conventionally inpainted skin image, thus indicating an improvement in extraction accuracy. The next step for this algorithm is to improve the compositing procedure by introducing more constraints on pixel label assignments and to develop a procedure for quantitative evaluation of the results.

As an application to rigid skin registration, a MJPM algorithm was described. The analysis showed that feature-based rigid registration by homography and random sample consensus (RANSAC) reliably identifies a local ROI marked by the convex hull of matched keypoints where RMSE of registration is sub-pixel. However, it was noted that the current rigid registration method performs unreliably on skin image pairs with slightly different microrelief appearance. In a proof-of-concept evaluation,

the microrelief junction point matching algorithm was shown to accurately match two sets of junction points within the region of interest (ROI) for a pair of skin images taken a few months apart. The matching result exhibited sub-pixel median matching error, a maximum of 2 pixels mean matching error, invariance to in-plane rotation and robustness to various local and global nonlinear distortion scales. As future work, research on improving microrelief structure extraction and predicting missing junction points is required for improving matching accuracy. Also, an evaluation of MJPM on larger datasets with an assortment of skin types and pathological conditions will lead to more conclusive results on the efficacy of this new method for skin image analysis and tracking.

In conclusion, nonrigid skin image registration makes it possible for skin images taken over a period of time, from different viewpoints or with induced stretching to be quantitatively analyzed by computers without the use of physical markers. Consequently, rigorous analysis for skin pathological conditions and skin features such as microrelief, hair and pigmentation can be realized. With the aid of reliable methods for hair detection and hair removal from skin images, skin image registration can be applied to a wider variety of skin images. Therefore, the set of computational tools described in this thesis paves the way for new quantitative-based insights in overall skin health and monitoring.

Bibliography

[1] A.d.a.m. images. Available at http://www.adamimages.com/. Accessed: 2016-06-29.

[2] Dermweb dullrazor. Available at http://www.dermweb.com/dull_razor/. Accessed: 2016-06-29.

[3] Qaisar Abbas, Irene Fondón Garcia, M Emre Celebi, and Waqar Ahmad. A feature-preserving hair removal algorithm for dermoscopy images. *Skin Research and Technology*, 19(1):e27–e36, 2013.

[4] Christos Nikolaos E Anagnostopoulos, Dimitrios D Vergados, and Panagiotis Mintzias. Image registration of follow-up examinations in digital dermoscopy. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4. IEEE, 2013.

[5] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *NIPS*, volume 2, page 3, 2000.

[6] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):509–522, 2002.

[7] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is âĂIJnearest neighborâĂİ meaningful? In *International conference on database theory*, pages 217–235. Springer, 1999.

[8] G. Bradski. *Dr. Dobb's Journal of Software Tools*, 2000.

[9] Matthew Brown and David G Lowe. Recognising panoramas. In *ICCV*, volume 3, page 1218, 2003.

[10] Matthew Brown, Richard Szeliski, and Simon Winder. Multi-image matching using multi-scale oriented patches. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 510–517. IEEE, 2005.

[11] Barbara Buffoli, Fabio Rinaldi, Mauro Labanca, Elisabetta Sorbellini, Anna Trink, Elena Guanziroli, Rita Rezzani, and Luigi F Rodella. The human hair: from anatomy to physiology. *International journal of dermatology*, 53(3):331–341, 2014.

[12] Peter J Burt and Edward H Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236, 1983.

[13] Jean-François Cardoso. High-order contrasts for independent component analysis. *Neural computation*, 11(1):157–192, 1999.

[14] Tat-Jun Chin, Hanzi Wang, and David Suter. Robust fitting of multiple structures: The statistical learning approach. In *2009 IEEE 12th International Conference on Computer Vision*, pages 413–420. IEEE, 2009.

[15] Young-Hwan Choi, Yoon-Sik Tak, Seungmin Rho, and Eenjun Hwang. Accurate wrinkle representation scheme for skin age estimation. In *Multimedia and Ubiquitous Engineering (MUE), 2011 5th FTRA International Conference on*, pages 226–231. IEEE, 2011.

[16] Haili Chui and Anand Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2):114–141, 2003.

[17] Ondřej Chum, Tomáš Pajdla, and Peter Sturm. The geometric error for homographies. *Computer Vision and Image Understanding*, 97(1):86–102, 2005.

[18] Oana G Cula, Kristin J Dana, Frank P Murphy, and Babar K Rao. Skin texture modeling. *International Journal of Computer Vision*, 62(1-2):97–119, 2005.

[19] Beth A Drolet, Nancy B Esterly, and Ilona J Frieden. Hemangiomas in children. *New England Journal of Medicine*, 341(3):173–181, 1999.

[20] Xian Du, Brian W Anthony, and Nigel C Kojimoto. Grid-based matching for full-field large-area deformation measurement. *Optics and Lasers in Engineering*, 66:307–319, 2015.

[21] M Fiorese, E Peserico, and A Silletti. Virtualshave: automated hair removal from digital dermatoscopic images. In *Conference proceedings:... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, volume 2011, pages 5145–5148, 2010.

[22] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[23] George Elmer Forsythe, Cleve B Moler, and Michael A Malcolm. Computer methods for mathematical computations. 1977.

[24] David Fouhey. Multi-model estimation in the presence of outliers. Bachelor's thesis, Middlebury College, 2011.

[25] Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3):209–226, 1977.

[26] Luo Gang, Opas Chutatape, and Shankar M Krishnan. Detection and measurement of retinal vessels in fundus images using amplitude modified second-order gaussian filter. *IEEE transactions on Biomedical Engineering*, 49(2):168–172, 2002.

[27] AJ Gewirtzman, J-H Saurat, and RP Braun. An evaluation of dermoscopy fluids and application techniques. *British Journal of Dermatology*, 149(1):59–63, 2003.

[28] Robert M Haralick. A measure for circularity of digital figures. *IEEE Transactions on Systems, Man, and Cybernetics*, 4(SMC-4):394–396, 1974.

[29] Robert M Haralick, Stanley R Sternberg, and Xinhua Zhuang. Image analysis using mathematical morphology. *IEEE transactions on pattern analysis and machine intelligence*, (4):532–550, 1987.

[30] Martha R Harkey. Anatomy and physiology of hair. *Forensic Science International*, 63(1-3):9–18, 1993.

[31] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision.* Cambridge university press, 2003.

[32] Richard I Hartley. In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(6):580–593, 1997.

[33] Adam Huang, Shun-Yuen Kwan, Wen-Yu Chang, Min-Yin Liu, Min-Hsiu Chi, and Gwo-Shing Chen. A robust hair segmentation and removal approach for clinical images of skin lesions. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3315–3318. IEEE, 2013.

[34] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000.

[35] Takanori Igarashi, Ko Nishino, and Shree K Nayar. The appearance of human skin: A survey. *Foundations and Trends® in Computer Graphics and Vision*, 3(1):1–95, 2007.

[36] Hossam Isack and Yuri Boykov. Energy-based geometric multi-model fitting. *International journal of computer vision*, 97(2):123–147, 2012.

[37] Anil Jain, Arun Ross, and Salil Prabhakar. Fingerprint matching using minutiae and texture features. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 282–285. IEEE, 2001.

[38] Bing Jian and Baba C Vemuri. Robust point set registration using gaussian mixture models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(8):1633–1645, 2011.

[39] Xudong Jiang and Wei-Yun Yau. Fingerprint minutiae matching based on the local and global structures. In *Pattern recognition, 2000. Proceedings. 15th international conference on*, volume 2, pages 1038–1041. IEEE, 2000.

[40] Timor Kadir, Andrew Zisserman, and Michael Brady. An affine invariant salient region detector. In *Computer Vision-ECCV 2004*, pages 228–241. Springer, 2004.

[41] Yasushi Kanazawa and Hiroshi Kawakami. Detection of planar regions with uncalibrated stereo using distributions of feature points. In *BMVC*, pages 1–10. Citeseer, 2004.

[42] Kimia Kiani and Ahmad R Sharafat. E-shaver: An improved dullrazor ® for digitally removing dark and light-colored hairs in dermoscopic images. *Computers in biology and medicine*, 41(3):139–145, 2011.

[43] Kyungrok Kim, Young-Hwan Choi, and Eenjun Hwang. Wrinkle feature-based skin age estimation scheme. In *2009 IEEE International Conference on Multimedia and Expo*, pages 1222–1225. IEEE, 2009.

[44] Stefan Klein, Marius Staring, and Josien PW Pluim. Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing*, 16(12):2879–2890, 2007.

[45] Joost Koehoorn, André Sobiecki, Paulo Rauber, Andrei Jalba, and Alexandru Telea. Efficient and effective automated digital hair removal from dermoscopy images. *Mathematical Morphology-Theory and Applications*, 1(1), 2016.

[46] Joost Koehoorn, André C Sobiecki, Daniel Boda, Adriana Diaconeasa, Susan Doshi, Stephen Paisey, Andrei Jalba, and Alexandru Telea. Automated digital hair removal by threshold decomposition and morphological analysis. In *International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 15–26. Springer, 2015.

[47] Konstantin Korotkov and Rafael Garcia. Computerized analysis of pigmented skin lesions: a review. *Artificial intelligence in medicine*, 56(2):69–90, 2012.

[48] Ina A. Kundu. Imaging platforms for detecting and analyzing skin features and its stability - with applications in skin health and in using the skin as a body-relative position-encoding system. Master's thesis, Massachusetts Institute of Technology, 2015.

[49] Jan Kybic and Michael Unser. Fast parametric elastic image registration. *IEEE transactions on image processing*, 12(11):1427–1442, 2003.

[50] Der-Tsai Lee and Bruce J Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242, 1980.

[51] Ik-Hyun Lee and Tae-Sun Choi. Accurate registration using adaptive block processing for multispectral images. *IEEE transactions on circuits and systems for video technology*, 23(9):1491–1501, 2013.

[52] Seungyong Lee, George Wolberg, and Sung Yong Shin. Scattered data interpolation with multilevel b-splines. *IEEE transactions on visualization and computer graphics*, 3(3):228–244, 1997.

[53] Tim Lee, Vincent Ng, Richard Gallagher, Andrew Coldman, and David McLean. Dullrazor®: A software approach to hair removal from images. *Computers in biology and medicine*, 27(6):533–543, 1997.

[54] Dong Li. *Multi-view, high-resolution face image analysis*. PhD thesis, The Hong Kong Polytechnic University, 2014.

[55] Li Li, Sophie Mac-Mary, David Marsaut, Jean Marie Sainthillier, Stéphanie Nouveau, Tijani Gharbi, Olivier de Lacharriere, and Philippe Humbert. Age-related changes in skin topography and microcirculation. *Archives of dermatological research*, 297(9):412–416, 2006.

[56] Chengjun Liu and Jian Yang. Ica color space for pattern recognition. *IEEE Transactions on Neural Networks*, 20(2):248–257, 2009.

[57] Laura Liu, David Zhang, and Jane You. Detecting wide lines using isotropic nonlinear filtering. *IEEE Transactions on image processing*, 16(6):1584–1595, 2007.

[58] Philippe Longere, Xuemei Zhang, Peter B Delahunt, and David H Brainard. Perceptual assessment of demosaicing algorithm performance. *Proceedings of the IEEE*, 90(1):123–132, 2002.

[59] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[60] S Madan, KJ Dana, and GO Cula. Multimodal and time-lapse skin registration. *Skin Research and Technology*, 21(3):319–326, 2015.

[61] Siddharth K Madan, Kristin J Dana, and Oana G Cula. Quasiconvex alignment of multimodal skin images for quantitative dermatology. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 117–124. IEEE, 2009.

169

[62] Ilias Maglogiannis. Automated segmentation and registration of dermatological images. *Journal of Mathematical Modelling and Algorithms*, 2(3):277–294, 2003.

[63] Ilias Maglogiannis and Kostantinos Delibasis. Hair removal on dermoscopy images. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2960–2963. IEEE, 2015.

[64] Jamaluddin Mahmud, Cathy A Holt, and Sam L Evans. An innovative application of a small-scale motion analysis technique to quantify human skin deformation in vivo. *Journal of biomechanics*, 43(5):1002–1006, 2010.

[65] Ashfaq A Marghoob, Lucinda D Swindle, Claudia ZM Moricz, Fitzgeraldo A Sanchez Negron, Bill Slue, Allan C Halpern, and Alfred W Kopf. Instruments and new technologies for the in vivo diagnosis of melanoma. *Journal of the American Academy of Dermatology*, 49(5):777–797, 2003.

[66] Stephen R Marschner, Stephen H Westin, Eric PF Lafortune, Kenneth E Torrance, and Donald P Greenberg. Reflectance measurements of human skin. 1999.

[67] Scott W Menzies, Alex Gutenev, Michelle Avramidis, Andrew Batrac, and William H McCarthy. Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesions. *Archives of dermatology*, 137(12):1583–1589, 2001.

[68] Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1):63–86, 2004.

[69] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10):1615–1630, 2005.

[70] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2):43–72, 2005.

[71] E Montin, E Cutrì, G Spadola, A Testori, G Pennati, and L Mainardi. Tuning of a deformable image registration procedure for skin component mechanical properties assessment. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6305–6308. IEEE, 2015.

[72] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

[73] Andriy Myronenko, Xubo Song, and Miguel A Carreira-Perpinán. Non-rigid point set registration: Coherent point drift. In *Advances in Neural Information Processing Systems*, pages 1009–1016, 2006.

[74] Nhi Hoang Nguyen. A hybrid approach to segmenting hair in dermoscopic images using a universal kernel. Master's thesis, Simon Fraser University, 2009.

[75] United States. National Bureau of Standards and Fred Edwin Nicodemus. *Geometrical considerations and nomenclature for reflectance*, volume 160. US Department of Commerce, National Bureau of Standards, 1977.

[76] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *Automatica*, 11(285-296):23–27, 1975.

[77] SA Pavlopoulos. New hybrid stochastic-deterministic technique for fast registration of dermatological images. *Medical and Biological Engineering and Computing*, 42(6):777–786, 2004.

[78] Dzung L Pham, Chenyang Xu, and Jerry L Prince. Current methods in medical image segmentation 1. *Annual review of biomedical engineering*, 2(1):315–337, 2000.

[79] Gérald E Piérard, Isabelle Uhoda, and Claudine Piérard-Franchimont. From skin microrelief to wrinkles. an area ripe for investigation. *Journal of cosmetic dermatology*, 2(1):21–28, 2003.

[80] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[81] Denis Rutovitz. Pattern recognition. *Journal of the Royal Statistical Society. Series A (General)*, 129(4):504–530, 1966.

[82] Frederik Schaffalitzky and Andrew Zisserman. Multi-view matching for unordered image sets, or âĂIJhow do i organize my holiday snaps?âĂİ. In *Computer VisionâĂŤECCV 2002*, pages 414–431. Springer, 2002.

[83] Philippe Schmid-Saugeon, Joël Guillodb, and Jean-Philippe Thirana. Towards a computer-aided diagnosis system for pigmented skin lesions. *Computerized Medical Imaging and Graphics*, 27(1):65–78, 2003.

[84] Clayton Scott and Robert Nowak. Robust contour matching via the order-preserving assignment problem. *IEEE Transactions on Image Processing*, 15(7):1831–1838, 2006.

[85] Michele Setaro and Adele Sparavigna. Irregularity skin index (isi): a tool to evaluate skin surface texture. *Skin Research and Technology*, 7(3):159–163, 2001.

[86] SO Skrovseth, Thomas R Schopf, Kevin Thon, Maciel Zortea, Marc Geilhufe, Kajsa Mollersen, Herbert M Kirchesch, and Fred Godtliebsen. A computer aided diagnostic system for malignant melanomas. In *2010 3rd International*

Symposium on Applied Sciences in Biomedical and Communication Technologies (ISABEL 2010), 2010.

[87] Noah Snavely, Steven M Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.

[88] Michal Sofka and Charles V Stewart. Retinal vessel centerline extraction using multiscale matched filters, confidence and edge measures. *IEEE transactions on medical imaging*, 25(12):1531–1546, 2006.

[89] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.

[90] Carsten Steger. An unbiased detector of curvilinear structures. *IEEE Transactions on pattern analysis and machine intelligence*, 20(2):113–125, 1998.

[91] Stanley R Sternberg. Grayscale morphology. *Computer vision, graphics, and image processing*, 35(3):333–355, 1986.

[92] Shih-Yu Sun. *Ultrasound probe localization by tracking skin features*. PhD thesis, Massachusetts Institute of Technology, 2014.

[93] Alexandru Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004.

[94] J-P Thirion. Image matching as a diffusion process: an analogy with maxwell's demons. *Medical image analysis*, 2(3):243–260, 1998.

[95] Marius Tico and Pauli Kuosmanen. Fingerprint matching using an orientation-based minutia descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):1009–1014, 2003.

[96] Roberto Toldo and Andrea Fusiello. Robust multiple structures estimation with j-linkage. In *European conference on computer vision*, pages 537–547. Springer, 2008.

[97] Mohammad Taghi Bahreyni Toossi, Hamid Reza Pourreza, Hoda Zare, Mohamad-Hoseyn Sigari, Pouran Layegh, and Abbas Azimi. An effective hair removal algorithm for dermoscopy images. *Skin Research and Technology*, 19(3):230–235, 2013.

[98] Luc Van Gool, Theo Moons, and Dorin Ungureanu. Affine/photometric invariants for planar intensity patterns. In *Computer Vision-ECCV'96*, pages 642–651. Springer, 1996.

[99] Andrea Vedaldi and Stefano Soatto. Features for recognition: Viewpoint invariance for non-planar scenes. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1474–1481. IEEE, 2005.

172

[100] Etienne Vincent and Robert Laganiére. Detecting planar homographies in an image pair. In *Proceedings of the 2nd International Symposium on Image and Signal Processing and Analysis*, pages 182–187, 2001.

[101] Luc Vincent and Pierre Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE transactions on pattern analysis and machine intelligence*, 13(6):583–598, 1991.

[102] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[103] Feng-Ying Xie, Shi-Yin Qin, Zhi-Guo Jiang, and Ru-Song Meng. Pde-based unsupervised repair of hair-occluded information in dermoscopy images of melanoma. *Computerized Medical Imaging and Graphics*, 33(4):275–282, 2009.

[104] Fengying Xie, Yang Li, Rusong Meng, and Zhiguo Jiang. No-reference hair occlusion assessment for dermoscopy images based on distribution feature. *Computers in biology and medicine*, 59:106–115, 2015.

[105] Neil Yager and Adnan Amin. Fingerprint verification based on minutiae features: a review. *Pattern Analysis and Applications*, 7(1):94–113, 2004.

[106] Sebastian Zambanini, Georg Langs, Robert Sablatnig, and Harald Maier. Automatic robust registration of cutaneous hemangiomas for follow-up examinations 1). 2007.

[107] Wei Zhang and Jana Košecká. Nonparametric estimation of multiple structures with outliers. In *Dynamical Vision*, pages 60–74. Springer, 2007.

[108] Feng Zhao and Xiaoou Tang. Preprocessing and postprocessing for skeleton-based fingerprint minutiae extraction. *Pattern Recognition*, 40(4):1270–1281, 2007.

[109] Howard Zhou, Mei Chen, Richard Gass, James M Rehg, Laura Ferris, Jonhan Ho, and Laura Drogowski. Feature-preserving artifact removal from dermoscopy images. In *Medical Imaging*, pages 69141B–69141B. International Society for Optics and Photonics, 2008.

[110] Barbara Zitova and Jan Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.

[111] Yaobin Zou, Enmin Song, and Renchao Jin. Age-dependent changes in skin surface assessed by a novel two-dimensional image analysis. *Skin research and technology*, 15(4):399–406, 2009.

[112] Marco Zuliani. Ransac for dummies. *Vision Research Lab, University of California, Santa Barbara*, 2009.

[113] Marco Zuliani, Charles S Kenney, and BS Manjunath. The multiransac algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–153. IEEE, 2005.